PATENT ABSTRACTS OF JAPAN

(11)Publication number : **11-065590**

(43)Date of publication of application : **09.03.1999**

(51)Int.Cl. **G10L 3/00**
G10L 3/00
H04Q 7/38
H04M 1/274
H04M 3/42

(21)Application number : **09-228567**

(22)Date of filing : **25.08.1997**

(71)Applicant : **NEC CORP**

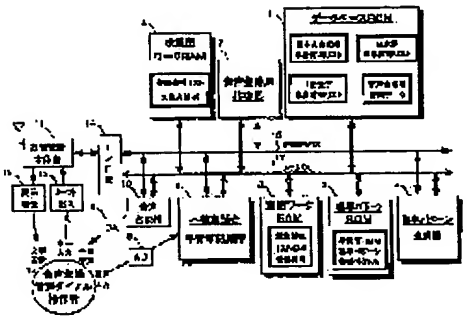
(72)Inventor : **TABEI KAZUHIKO**

(54) VOICE RECOGNITION DIALING DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce cumbersome key operations for registering names and corresponding telephone numbers on a portable telephone by outputting the recognition result of the names and the telephone numbers by the recognition processing, which uses the monosyllable code data stream of the registered names or the telephone numbers with respect to input voices and standard pattern data in half syllable units.

SOLUTION: A voice registration means, such as a general speaker half syllable voice recognizer 6 registers the uttered voices of names and telephone numbers through the voice input of the object names and the corresponding telephone numbers for a voice recognition dialing. The dialing voice means of the recognizer 6 conductors a dialing, using the monosyllablic code data column of the names and the telephone number which have been registered beforehand from the voice analog signals of the names and the telephone numbers. The recognizer 6 conducts the recognition process, which uses the monosyllablic code data column of the names and the telephone numbers that have been registered with respect to input voice and the standard pattern data in half syllable nits. Then, the name of the telephone number monosyllablic code data column of the candidate having a small cumulative distance value is outputted as the recognized result.



LEGAL STATUS

[Date of request for examination]	25.08.1997
[Date of sending the examiner's decision of rejection]	10.05.2000
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3447521

[Date of registration]	04.07.2003
[Number of appeal against examiner's decision of rejection]	2000-08467
[Date of requesting appeal against examiner's decision of rejection]	08.06.2000
[Date of extinction of right]	

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－65590

(43)公開日　平成11年(1999)３月９日

(51)Int.Cl.⁶		識別記号	FI		
Ｇ１０Ｌ	3/00	５５１	Ｇ１０Ｌ	3/00	５５１Ａ
		５２１			５２１
Ｈ０４Ｑ	7/38		Ｈ０４Ｍ	1/274	
Ｈ０４Ｍ	1/274			3/42	Ｐ
	3/42		Ｈ０４Ｂ	7/26	１０９Ｑ

審査請求　有　　請求項の数９　ＯＬ　（全 25 頁）

(21)出願番号　特願平9－228567

(22)出願日　平成９年(1997)８月25日

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号

(72)発明者　田部井　和彦
東京都港区芝五丁目７番１号　日本電気株式会社内

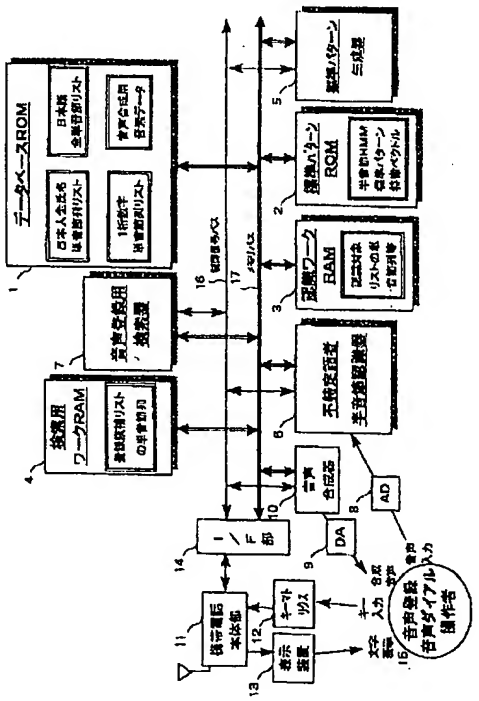
(74)代理人　弁理士　若林　忠　（外４名）

(54)【発明の名称】　音声認識ダイアル装置

(57)【要約】

【課題】　携帯電話機等における氏名と電話番号登録操作、およびダイアル操作等のキー操作における煩雑さを低減する音声認識ダイアル装置を提供する。

【解決手段】　日本人全氏名のデータベースを単音節コード単位（ひらがな文字コード；ＡＳＣＩＩあるいはＳＪＩＳ等）であらかじめデータベースとしてＲＯＭ等に格納しておき、これと単音節単位で認識処理可能な不特定話者半音節音声認識器により、氏名および電話番号を単音節単位で離散発声した音声信号から自動登録させることによる音声登録機能と、登録された氏名および電話番号の単音節コードを認識対象とした不特定話者半音節音声認識器による認識処理により、音声ダイアルを行う機能とにより、キー操作を低減できる。



1

【特許請求の範囲】

【請求項１】　氏名と電話番号を発声した音声のＰＣＭ信号から音声認識処理により単音節コードデータ列を取得し、音声認識ダイアル用の対象氏名と電話番号を音声入力で登録する音声登録手段と、

氏名又は電話番号の音声アナログ信号からあらかじめ登録済みの氏名と電話番号の単音節コードデータ列を用いてダイアルをする音声ダイアル手段と、

入力音声に対する、登録済み氏名又は電話番号の単音節コードデータ列と半音節単位の標準パターンデータとを用いた認識処理により、累積距離値の小さい候補の氏名又は電話番号の単音節コードデータ列を認識結果として出力する不特定話者半音節音声認識器とを備えている音声認識ダイアル装置。

【請求項２】　前記音声登録手段は、

音声アナログ信号をＰＣＭ信号へ変換するＡＤコンバータと、

前記ＰＣＭ信号を前記音声アナログ信号へ変換するＤＡコンバータと、

日本語の１音節を表す単音節コードデータを受信しひらがなと数字と漢字の表記文字を表示する表示手段と、

前記単音節コードデータを受信し音声ＰＣＭデータに変換しＤＡコンバータに出力する音声応答手段と、

日本人氏名と日本語の全単音節と数字１桁の各々について発音表記の単音節列を単音節コードデータで格納したデータベースＲＯＭと、

該データベースＲＯＭ内の氏名項目又は単音節項目又は１桁数字項目の何れかの項目に属する１氏名又は１単音節又は１桁数字の何れかの１リストに対する単音節コードデータ列を読出し認識ワークＲＡＭへ格納する手段と、

単音節単位又は数字１桁単位に発声された前記音声アナログ信号をＡＤ変換した前記ＰＣＭ信号を前記不特定話者半音節音声認識器へ入力し、前記単音節単位又は数字１桁単位の認識処理を行い第１候補あるいは複数候補の認識結果を得る手段と、

前記単音節単位又は数字１桁単位の音声認識結果情報と前記データベースＲＯＭ内の氏名項目又は１桁数字項目の単音節コードデータ列との両方の情報を用いて発声内容に最も近いと推測される氏名あるいは数字１桁以上の単音節コードデータ列を検索し出力する音声登録用検索器と、

検索された氏名又は電話番号の単音節コードデータ列を各氏名と電話番号を対応付けて複数の該各氏名と電話番号リストを認識ワークＲＡＭへ蓄積格納する手段とを備えている請求項１に記載の音声認識ダイアル装置。

【請求項３】　前記音声登録用検索器は、

前記不特定話者半音節認識器から前記単音節単位又は数字１桁単位の音声認識結果情報を累積距離値と共に情報受信し、前記データベースＲＯＭに格納されている氏名

2

項目又は１単音節又は１桁数字項目の何れかの項目に属する１氏名又は１単音節又は１桁数字の何れかの１リスト内で、１単音節単位に発声された音声信号から最も距離の近い単音節コードデータ列を検索し認識ワークＲＡＭへ格納する手段を有する請求項２に記載の音声認識ダイアル装置。

【請求項４】　前記音声ダイアル手段は、

ダイアルの宛先の氏名又は電話番号に対応する発声した音声アナログ信号をＡＤコンバータで変換したＰＣＭ信号を前記不特定話者半音節音声認識器へ入力し、認識ワークＲＡＭ内の登録済み氏名又は電話番号リストに対して前記発声した音声アナログ信号に最も近いと推測される認識結果を単音節コードデータ列で取得する手段と、

認識結果が氏名の場合は対応付けされた電話番号の単音節コードデータ列を前記認識ワークＲＡＭ内から検索して出力する手段と、

前記電話番号の単音節コードデータ列から電話端末本体への電話発呼信号へ変換する手段とを有する請求項１に記載の音声認識ダイアル装置。

【請求項５】　前記不特定話者半音節音声認識器は、

認識ワークＲＡＭ又はデータベースＲＯＭに格納されている氏名又は電話番号の単音節単位の単音節コードデータ列に対し半音節単位の半音節コードデータ列へ変換する手段と、

氏名又は電話番号の前記半音節コードデータ列に対して標準パターン生成器の不特定話者半音節音声認識装置用の標準パターンデータ生成により得られた標準パターンを格納した標準パターンＲＯＭ内からどの半音節単位の隠れマルコフモデルが含まれているかを調べ、さらに前記隠れマルコフモデル状態の連結を示す半音節隠れマルコフモデル状態コードデータ列へ変換する手段と、

氏名又は電話番号の前記半音節隠れマルコフモデル状態コードデータ列を氏名－電話番号の関係で対応付けし、さらに他の氏名－電話番号リストと識別可能なように番号付きリストに変換して前記認識ワークＲＡＭへ格納する手段と、

音声アナログ信号をＡＤ変換した音声ＰＣＭ信号からフレーム単位の入力特徴ベクトルを抽出する音声分析特徴抽出器と、

抽出された前記入力特徴ベクトルを前記認識ワークＲＡＭへ格納する手段と、

前記入力特徴ベクトルと、標準半音節隠れマルコフモデルパターンＲＯＭに格納されている全半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を算出する状態距離計算器と、

算出された前記状態距離値に番号付けして認識ワークＲＡＭに格納する手段と、

前記認識ワークＲＡＭに格納された認識対象の各氏名又は電話番号に対する前記半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と前記標

準パターンＲＯＭ内に格納されている状態間遷移距離値とを用いてフレーム同期Ｖｉｔｅｒｂｉアルゴリズムにより入力音声時間長分の全フレームに対する累積距離値を算出する累積状態距離計算器と、

最も前記累積距離値の小さい第１候補又は第１～第Ｎ候補の氏名又は電話番号に対する単音節コードデータ列を認識結果として出力する手段とを有する請求項１から請求項４の何れか１項に記載の音声認識ダイアル装置。

【請求項６】　前記標準パターン生成器は、

統計的に必要とされる人数分の多数話者の音声アナログ信号をＡＤ変換したＰＣＭ信号を、波形表示あるいは試聴等により所定数の種類の単音節単位のＰＣＭ信号へ分割する手段と、

前記単音節単位に分割された全てのＰＣＭ信号をバッファリングし、波形表示あるいは試聴等により所定数の種類の半音節単位のＰＣＭデータ信号へ分割する手段と、

前記半音節単位に分割された全ての各ＰＣＭ信号をフレーム単位に分割する手段と、

前記フレーム単位に分割された全てのＰＣＭ信号に対して特徴ベクトルを抽出する音声分析特徴抽出器と、

状態数が所定の個数の隠れマルコフモデルにおいて前記所定の個数の状態出力確率関数と前記所定の個数の２倍個数分の状態遷移確率のパラメータを初期値設定する手段と、

前記状態出力確率関数を初期設定する際に多次元正規分布確率密度関数を用いると共に、母数として平均ベクトルおよび共分散行列の各成分を初期値設定する手段と、

所定の種類分の半音節毎に得られた前記統計的に必要とされる人数分のフレーム単位の前記特徴ベクトルから、各半音節毎に前記統計的に必要とされる人数分の特徴ベクトルサンプルとして整理して、Ｆｏｒｗａｒｄ－Ｂａｃｋｗａｒｄアルゴリズムという反復的手法により前記所定の個数の平均ベクトルおよび共分散行列の各成分値と前記所定の個数の２倍個数分の状態遷移確率とを得る手段と、

所定の種類分の半音節毎に得られた隠れマルコフモデルのパラメータ群である多次元正規分布確率密度関数の平均ベクトルと共分散行列と状態遷移確率とのパラメータ値を標準パターンデータとして標準パターンＲＯＭへ格納する手段とを有する請求項５に記載の音声認識ダイアル装置。

【請求項７】　前記音声分析特徴抽出器は、

入力音声のＰＣＭ信号をフレーム分割したフレームＰＣＭ信号をプリエンファシス処理することにより高周波数帯域を強調する手段と、

前記プリエンファシス処理済みフレームＰＣＭ信号に対し窓処理することによりこの後のＦＦＴ処理のためのフレーム境界のスムージング処理をする手段と、

前記窓処理後のフレームＰＣＭ信号をＮ次ＦＦＴ変換処理することにより線形周波数軸上のＮ次複素係数ベクトルへ変換する手段と、

該Ｎ次複素係数ベクトルから複素数の絶対値計算によりＮ次振幅係数ベクトルへ変換する手段と、

該Ｎ次振幅係数ベクトルに対し対数演算を施してＮ次対数振幅係数ベクトルを算出する手段と、

該Ｎ次対数振幅係数ベクトルに対し時間軸上への逆離散余弦変換によりＰ次ケプストラムベクトルを算出する手段と、

該Ｐ次ケプストラムベクトルの高時間成分を抑圧する処理により、声道特性とピッチ特性を分離し声道特性のみを抽出したＰ次声道特性ケプストラムベクトルを算出する手段と、

該Ｐ次声道特性ケプストラムベクトルに対し周波数軸上への離散余弦変換を行い、線形周波数軸上におけるＮ次声道特性対数振幅係数ベクトルに変換する手段と、

該Ｎ次声道特性対数振幅係数ベクトルに対しメル周波数軸上における等分割点上のスペクトル成分へ補間あるいはスムージング処理を施したベクトル成分を算出することにより人間の聴覚周波数分解能特性に合わせたＮ次声道特性対数振幅メル尺度係数ベクトルへ変換する手段と、

該Ｎ次声道特性対数振幅メル尺度係数ベクトルに対し時間軸上への逆離散余弦変換によりＱ次メルケプストラムベクトルへ変換する手段とを有する請求項５又は請求項６に記載の音声認識ダイアル装置。

【請求項８】　前記状態距離計算器は、

入力音声のアナログ信号をＡＤ変換しさらにフレーム分割したフレームＰＣＭ信号から前記音声分析特徴抽出器により得られた入力特徴ベクトルと標準半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を所定の距離計算式により算出する手段を有する請求項５に記載の音声認識ダイアル装置。

【請求項９】　前記累積距離計算器は、

認識対象の氏名又は電話番号に対する半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と標準パターンＲＯＭ内に格納されている状態間遷移距離値とを用いてフレーム同期Ｖｉｔｅｒｂｉアルゴリズムにより入力音声の全フレームに対する累積距離値を算出する手段を有する請求項５に記載の音声認識ダイアル装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、音声認識ダイアル装置に関する。

【０００２】

【従来の技術】従来では、認識単語登録の作業には使用者のキー入力装置への操作が必要であった。例えば、氏名と電話番号の１人分の登録を実行しようとするには、使用者が［人名の平仮名文字数］＋［電話番号桁数］分のキー操作入力が必要となる。

5

【０００３】携帯電話では小型化のために、パソコンのような全文字分の入力キーを持った入力装置ではなく、入力キーの数量が限られているため、複数文字分が重複されて割り当てられている。

【０００４】例えば、携帯電話の入力装置を例に、人名＝７文字、電話番号＝１０桁を登録しようとする場合を考える。

【０００５】携帯電話では、ア行＝１キー、カ行＝１キー～ラ行＝１キー、ワヲン＝１キーという形でキーが割り当てられており、７文字を登録する場合最短で１×７＝７回のキー入力、最長で５×７＝３５回のキー入力、平均で２．５×７＝１７．５回のキー入力となる。

【０００６】数字は全部がキーに割り当てられており、操作モード変更することで入力できるため、１＋１０回のキー入力となる。

【０００７】以上の例では、平均操作時間を１キー入力当たり＝２秒とすると２秒×（１７．５＋１０）回＝５５秒となる。

【０００８】

【発明が解決しようとする課題】上述した従来の技術では、音声認識ダイアル用の氏名と電話番号の登録操作時間が長く、また操作間違いも少なくない。

【０００９】また音声を用いた従来技術に特開平２－１３５８４７号公報に開示された音声応答認識自動ダイアル電話機があるが、この従来技術では、音声入力信号を認識し文字データに、変換する機能を用いているが、実現手段が明記されていない。

【００１０】本発明の目的は、携帯電話機等における氏名と電話番号登録操作、およびダイアル操作等のキー操作における煩雑さを低減する音声認識ダイアル装置を提供することにある。

【００１１】

【課題を解決するための手段】本発明の音声認識ダイアル装置は、氏名と電話番号を発声した音声のＰＣＭ信号から音声認識処理により単音節コードデータ列を取得し、音声認識ダイアル用の対象氏名と電話番号を音声入力で登録する音声登録手段と、氏名又は電話番号の音声アナログ信号からあらかじめ登録済みの氏名と電話番号の単音節コードデータ列を用いてダイアルをする音声ダイアル手段と、入力音声に対する、登録済み氏名又は電話番号の単音節コードデータ列と半音節単位の標準パターンデータとを用いた認識処理により、累積距離値の小さい候補の氏名又は電話番号の単音節コードデータ列を認識結果として出力する不特定話者半音節音声認識器とを備えている。

【００１２】また、音声登録手段は、音声アナログ信号をＰＣＭ信号へ変換するＡＤコンバータと、ＰＣＭ信号を音声アナログ信号へ変換するＤＡコンバータと、日本語の１音節を表す単音節コードデータを受信しひらがなと数字と漢字の表記文字を表示する表示手段と、単音節

6

コードデータを受信し音声ＰＣＭデータに変換しＤＡコンバータに出力する音声応答手段と、日本人氏名と日本語の全単音節と数字１桁の各々について発音表記の単音節列を単音節コードデータで格納したデータベースＲＯＭと、データベースＲＯＭ内の氏名項目又は単音節項目又は１桁数字項目の何れかの項目に属する１氏名又は１単音節又は１桁数字の何れかの１リストに対する単音節コードデータ列を読出し認識ワークＲＡＭへ格納する手段と、単音節単位又は数字１桁単位に発声された音声アナログ信号をＡＤ変換したＰＣＭ信号を不特定話者半音節音声認識器へ入力し、単音節単位又は数字１桁単位の認識処理を行い第１候補あるいは複数候補の認識結果を得る手段と、単音節単位又は数字１桁単位の音声認識結果情報とデータベースＲＯＭ内の氏名項目又は１桁数字項目の単音節コードデータ列との両方の情報を用いて発声内容に最も近いと推測される氏名あるいは数字１桁以上の単音節コードデータ列を検索し出力する音声登録用検索器と、検索された氏名又は電話番号の単音節コードデータ列を各氏名と電話番号を対応付けて複数の各氏名と電話番号リストを認識ワークＲＡＭへ蓄積格納する手段とを備えていてもよい。

【００１３】また、音声登録用検索器は、不特定話者半音節認識器から単音節単位又は数字１桁単位の音声認識結果情報を累積距離値と共に情報受信し、データベースＲＯＭに格納されている氏名項目又は１単音節又は１桁数字項目の何れかの項目に属する１氏名又は１単音節又は１桁数字の何れかの１リスト内で、１単音節単位に発声された音声信号から最も距離の近い単音節コードデータ列を検索し認識ワークＲＡＭへ格納する手段を有してもよい。

【００１４】また、音声ダイアル手段は、ダイアルの宛先の氏名又は電話番号に対応する発声した音声アナログ信号をＡＤコンバータで変換したＰＣＭ信号を不特定話者半音節音声認識器へ入力し、認識ワークＲＡＭ内の登録済み氏名又は電話番号リストに対して発声した音声アナログ信号に最も近いと推測される認識結果を単音節コードデータ列で取得する手段と、認識結果が氏名の場合は対応付けされた電話番号の単音節コードデータ列を認識ワークＲＡＭ内から検索して出力する手段と、電話番号の単音節コードデータ列から電話端末本体への電話発呼信号へ変換する手段とを有してもよい。

【００１５】また、不特定話者半音節音声認識器は、認識ワークＲＡＭ又はデータベースＲＯＭに格納されている氏名又は電話番号の単音節単位の単音節コードデータ列に対し半音節単位の半音節コードデータ列へ変換する手段と、氏名又は電話番号の半音節コードデータ列に対して標準パターン生成器の不特定話者半音節音声認識装置用の標準パターンデータ生成により得られた標準パターンを格納した標準パターンＲＯＭ内からどの半音節単位の隠れマルコフモデルが含まれているかを調べ、さら

に隠れマルコフモデル状態の連結を示す半音節隠れマルコフモデル状態コードデータ列へ変換する手段と、氏名又は電話番号の半音節隠れマルコフモデル状態コードデータ列を氏名－電話番号の関係で対応付けし、さらに他の氏名－電話番号リストと識別可能なように番号付きリストに変換して認識ワークＲＡＭへ格納する手段と、音声アナログ信号をＡＤ変換した音声ＰＣＭ信号からフレーム単位の入力特徴ベクトルを抽出する音声分析特徴抽出器と、抽出された入力特徴ベクトルを認識ワークＲＡＭへ格納する手段と、入力特徴ベクトルと、標準半音節隠れマルコフモデルパターンＲＯＭに格納されている全半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を算出する状態距離計算器と、算出された状態距離値に番号付けして認識ワークＲＡＭに格納する手段と、認識ワークＲＡＭに格納された認識対象の各氏名又は電話番号に対する半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と標準パターンＲＯＭ内に格納されている状態間遷移距離値とを用いてフレーム同期Ｖｉｔｅｒｂｉアルゴリズムにより入力音声時間長分の全フレームに対する累積距離値を算出する累積状態距離計算器と、最も累積距離値の小さい第１候補又は第１～第Ｎ候補の氏名又は電話番号に対する単音節コードデータ列を認識結果として出力する手段とを有してもよい。

【００１６】また、標準パターン生成器は、統計的に必要とされる人数分の多数話者の音声アナログ信号をＡＤ変換したＰＣＭ信号を、波形表示あるいは試聴等により所定数の種類の単音節単位のＰＣＭ信号へ分割する手段と、単音節単位に分割された全てのＰＣＭ信号をバッファリングし、波形表示あるいは試聴等により所定数の種類の半音節単位のＰＣＭデータ信号へ分割する手段と、半音節単位に分割された全ての各ＰＣＭ信号をフレーム単位に分割する手段と、フレーム単位に分割された全てのＰＣＭ信号に対して特徴ベクトルを抽出する音声分析特徴抽出器と、状態数が所定の個数の隠れマルコフモデルにおいて所定の個数の状態出力確率関数と所定の個数の２倍個数分の状態遷移確率のパラメータを初期値設定する手段と、状態出力確率関数を初期設定する際に多次元正規分布確率密度関数を用いると共に、母数として平均ベクトルおよび共分散行列の各成分を初期値設定する手段と、所定の種類分の半音節毎に得られた統計的に必要とされる人数分のフレーム単位の特徴ベクトルから、各半音節毎に統計的に必要とされる人数分の特徴ベクトルサンプルとして整理して、Ｆｏｒｗａｒｄ－Ｂａｃｋｗａｒｄアルゴリズムという反復的手法により所定の個数の平均ベクトルおよび共分散行列の各成分値と所定の個数の２倍個数分の状態遷移確率とを得る手段と、所定の種類分の半音節毎に得られた隠れマルコフモデルのパラメータ群である多次元正規分布確率密度関数の平均ベクトルと共分散行列と状態遷移確率とのパラメータ値を

標準パターンデータとして標準パターンＲＯＭへ格納する手段とを有してもよい。

【００１７】さらに、音声分析特徴抽出器は、入力音声のＰＣＭ信号をフレーム分割したフレームＰＣＭ信号をプリエンファシス処理することにより高周波数帯域を強調する手段と、プリエンファシス処理済みフレームＰＣＭ信号に対し窓処理することによりこの後のＦＦＴ処理のためのフレーム境界のスムージング処理をする手段と、窓処理後のフレームＰＣＭ信号をＮ次ＦＦＴ変換処理することにより線形周波数軸上のＮ次複素係数ベクトルへ変換する手段と、Ｎ次複素係数ベクトルから複素数の絶対値計算によりＮ次振幅係数ベクトルへ変換する手段と、Ｎ次振幅係数ベクトルに対し対数演算を施してＮ次対数振幅係数ベクトルを算出する手段と、Ｎ次対数振幅係数ベクトルに対し時間軸上への逆離散余弦変換によりＰ次ケプストラムベクトルを算出する手段と、Ｐ次ケプストラムベクトルの高時間成分を抑圧する処理により、声道特性とピッチ特性を分離し声道特性のみを抽出したＰ次声道特性ケプストラムベクトルを算出する手段と、Ｐ次声道特性ケプストラムベクトルに対し周波数軸上への離散余弦変換を行い、線形周波数軸上におけるＮ次声道特性対数振幅係数ベクトルに変換する手段と、Ｎ次声道特性対数振幅係数ベクトルに対しメル周波数軸上における等分割点上のスペクトル成分へ補間あるいはスムージング処理を施したベクトル成分を算出することにより人間の聴覚周波数分解能特性に合わせたＮ次声道特性対数振幅メル尺度係数ベクトルへ変換する手段と、Ｎ次声道特性対数振幅メル尺度係数ベクトルに対し時間軸上への逆離散余弦変換によりＱ次メルケプストラムベクトルへ変換する手段とを有してもよい。

【００１８】さらに、状態距離計算器は、入力音声のアナログ信号をＡＤ変換しさらにフレーム分割したフレームＰＣＭ信号から音声分析特徴抽出器により得られた入力特徴ベクトルと標準半音節隠れマルコフモデルの全状態の標準特徴ベクトルとの状態距離値を所定の距離計算式により算出する手段を有してもよい。

【００１９】さらに、累積距離計算器は、認識対象の氏名又は電話番号に対する半音節隠れマルコフモデル状態コードデータ列の状態結合情報と各状態距離値と標準パターンＲＯＭ内に格納されている状態間遷移距離値とを用いてフレーム同期Ｖｉｔｅｒｂｉアルゴリズムにより入力音声の全フレームに対する累積距離値を算出する手段を有してもよい。

【００２０】従って、本発明により、音声認識ダイアル用の氏名と電話番号の登録操作時間が従来例と比べて短縮され、また操作間違いも少なくなる。

【００２１】また、従来技術の認識自動ダイアル電話機では、明記されていなかった音声から文字表示を行う音声認識処理について、この技術の実現手段を明記することで本発明の属する技術分野において実際に利用可能と

なる。

【００２２】

【発明の実施の形態】本発明の実施の形態について図面を参照して説明する。図１は本発明の実施の形態の全体構成を示すブロック図である。まず、記憶装置として２種類のＲＯＭ（データベースＲＯＭ１、標準パターンＲＯＭ２）と、２種類のＲＡＭ（認識ワークＲＡＭ３、検索ワークＲＡＭ４）とがある。

【００２３】また、機能ブロックとして標準パターン生成器５と、不特定話者半音節音声認識器６と、音声登録用検索器７と、ＡＤコンバータ８と、ＤＡコンバータ９と、音声合成器１０と、携帯電話本体部１１と、キーマトリクス１２と、表示装置１３と、Ｉ／Ｆ部１４とがある。以降では、音声登録動作と音声ダイアル動作に分けて実施内容を説明する。また、これに続き主要な機能ブロックの内部詳細動作を、標準パターン生成器５と、不特定話者半音節音声認識器６と、音声登録用検索器７の、各々について各記憶装置間との連係動作内容も含めて説明する。

【００２４】また、図１に示す機能ブロックに対する実施の形態としては、基本的に半導体集積回路と複合装置等により実現可能である。まずＲＯＭ１、２と、ＲＡＭ３、４と、ＡＤコンバータ８と、ＤＡコンバータ９などは、半導体集積回路となる。また、携帯電話本体１１と、キーマトリクス１２と、表示装置１３（これはＬＣＤ等）と、Ｉ／Ｆ部１４（これは携帯電話と拡張機器を接続するための拡張コネクタが利用可）は、複合装置となる。さらに、音声登録用検索器７と、不特定話者半音節音声認識器６と、音声合成器１０は、ＣＰＵあるいは音声信号処理を高速演算可能なＤＳＰと呼ばれるマイクロプログラム内蔵可能な半導体集積回路上におけるソフトウェアにより実現することが可能である。

【００２５】１．音声登録動作

まず、データベースＲＯＭ１に格納されている全ての単音節単位（ひらがな１文字、即ち５０音＋α（濁音、拗音等））の単音節コードデータ（ＡＳＣＩＩあるいはＪＩＳあるいはＳＪＩＳ等）を読み出し、更に各単音節コードデータを半音節コードデータ（半音節単位とは１音節をさらに半分に分割した音素単位をいい、例“た”の場合、ＴＡ→Ｔ－，－Ａに分割する）に変換して認識ワークＲＡＭ３へ格納する。なお、単音節→半音節への変換は後述の３．不特定話者半音節音声認識器に内蔵された機能である。

【００２６】次に登録する氏名と電話番号を離散単音節単位あるいは数字１桁単位に離散発声することにより、この登録氏名と登録電話番号のひらがな文字情報として単音節コードデータを、また数字情報として数字コードデータを以下に詳細に述べる（１）氏名登録動作および（２）電話番号登録動作により得ることができる。さらに、得られた氏名と電話番号の対応関連情報と、登録順の番号を付けて認識ワークＲＡＭ３へ格納しておく。これにより、一連の音声登録動作が終了する。以降の（１）および（２）では、氏名を“たかはし”、電話番号を“０３－１２３－４５６７”という例を用いて、音声登録動作の詳細内容について説明する。

【００２７】（１）氏名登録動作

図２は音声登録動作モードのフローチャートである。

【００２８】まず発声者は、“た”：“か”：“は”：“し”と単音節毎に無音間隔を入れて発声した音声アナログ信号をＡＤコンバータ８によりＰＣＭ信号へ変換する（Ｓ３）。これらの４つの単音節分の各ＰＣＭ信号に対して不特定話者半音節音声認識器６により、単音節単位の認識結果を得る（Ｓ４）。ここで認識結果は、近いものから第１～第５候補まで出力されるものとする。なお、複数の単音節認識結果候補を必要とする理由は、半音節等の音素単位の音声認識器では通常最小音素単位の認識精度が低く、例えば“た”と発声しても第１候補には同母音系の音節として、“あかさたなはまやらわ”（拗音、濁音等の同母音系も含む）のような認識結果が出力される確率が高いため、第１候補だけではほとんど正確な認識結果が得られないのである。そこで認識結果として複数候補を用いるならば真の発声音節の認識結果がこれらの複数候補に含まれる可能性が高くなり、さらに、次に述べる登録氏名決定のための音声登録用検索器７とデータベースＲＯＭ１の情報とから認識確度が上がることになる。

【００２９】さて、氏名例の“た”“か”“は”“し”という４音節の離散発声アナログ信号をＡＤ変換した各々のＰＣＭ信号に対して、全ての単音節を認識候補とした（実際には、データベースＲＯＭ１から認識ワークＲＡＭ３へ全ての単音節コードデータを転送しておく）不特定話者半音節音声認識器６により図３に示すように４音節×５候補分の認識結果を得る。

【００３０】これらの情報は４．音声登録用検索器で説明するように登録ワークＲＡＭへ格納される。

【００３１】次に音声登録用検索器７により、これらの認識結果とデータベースＲＯＭ１内の４音節に限定した（発声回数が４回のため）氏名リストから検索処理により登録氏名を決定する（Ｓ９）。

【００３２】なお、氏名の検索結果を複数候補とる場合には、表示機能あるいは音声合成機能等を用いて、キー入力等により最終的に利用者（＝発声者）選択させることも出来る（Ｓ１０）。

【００３３】（２）電話番号登録動作

図４は電話番号登録動作モードのフローチャートである。

【００３４】電話番号登録は、認識ワークＲＡＭ３に数字の単音節結合リスト「〝ぜ＋ろ（０）〟～〝き＋ゅ＋う（９）〟）をあらかじめロードしておき、電話番号の桁数分だけ数字１桁ずつ離散発声を行うことにより（Ｓ

１３）得られる認識結果を図５に示すように数字を直接表すコードに変換する（Ｓ１４）。なお、数字認識の場合は、単音節認識に比べ認識対象リストも１０程度であり、また２音節以上の認識の場合は、認識精度も上がるため、あらかじめ電話番号リストをデータベースＲＯＭ１に準備する必要がない。数字の結果を複数候補とる場合には、表示機能あるいは音声合成機能等を用いて、キー入力等により最終的に利用者（＝発声者）に選択させることも出来る（Ｓ１６）。

【００３５】なお、電話番号の登録の際は、携帯電話等においてはキー入力操作も選択可能にしておいてもよい。一般に数字のキー割り当ては、１桁ずつある場合が多いので氏名入力ほど煩雑さの程度が低いからである。

【００３６】２．音声ダイアル動作

図６は音声ダイアル動作モードのフローチャートである。

【００３７】ダイアルを始める前に氏名または電話番号の音声入力選択をキー入力等の指定により、認識ワークＲＡＭ３内に格納されている氏名あるいは電話番号のいずれを認識対象とするかの初期設定を行っておく（Ｓ２２、Ｓ２３、Ｓ３０）。

【００３８】次に氏名あるいは電話番号の発声を行い（Ｓ２４、Ｓ３１）、このアナログ音声信号をＡＤ変換したＰＣＭ信号に対して、不特定話者半音節音声認識器により認識処理を行う（Ｓ２５、Ｓ３２）。

【００３９】この認識結果は、初期設定において、氏名なのか、それとも電話番号であるかがわかっている。そのため電話番号の認識結果を得た場合は、その情報により電話番号の数字コード等からダイアル用の発呼信号へ変換してダイアルを実行できる（Ｓ３６、Ｓ３７）。氏名の認識結果を得た場合には、氏名と電話番号の関係付けされた情報から電話番号の数字コードを特定して同様にダイアルが可能となる（Ｓ２９、Ｓ３６、Ｓ３７）。なお、電話番号あるいは氏名の認識処理後にダイアルを実行する前に氏名の文字と電話番号の数字を表示したり、あるいは音声合成器により音を出力したりすることにより１回の確認手続きを入れたりする（Ｓ２７、Ｓ３４）ことで、より親和性のある音声ダイアル機能にすることも可能である。

【００４０】３．不特定話者半音節音声認識動作

（１）標準パターン生成処理

図７及び図８は標準パターン生成器の機能ブロック構成図である。

【００４１】図９は標準パターン生成器による標準パターン生成処理のフローチャートである。

【００４２】標準パターンの生成は、多数話者（Ｎmax 人とする）の発声音声サンプルから各半音節単位の隠れマルコフモデル（ＨＭＭ；Ｈｉｄｄｅｎ　Ｍａｒｃｏｖ　Ｍｏｄｅｌ）の確率パラメータを推定することになる。

【００４３】まず、全ての調音結合パターンを含んだバランス音素テキストを用意し、統計的に十分な多数の話者＝Ｎmax 人に発声させて（Ｓ３８）、ＡＤ変換し（Ｓ３９）、ＰＣＭ信号を一旦認識ワークＲＡＭ３へ格納しておく。次にＰＣＭデータを波形表示あるいはＤＡ変換することで目視あるいは試聴等の作業により、Ｈmax種類の単音節単位に区切り（Ｓ４０）、さらに前後の調音結合を考慮して分類したＩmax 種類の半音節単位毎に半音節ＰＣＭ信号を得る（Ｓ４１）。この処理の様子を図１１の例に示す。

【００４４】ここで“Ｋ－”の“－”は後方に音が続いていることを示しており、“－Ａ”の“－”は、前方の音に続くことを示している。

【００４５】以上までにおいて、半音節種類数×Ｎmax 人分のＰＣＭ信号サンプルが準備できたことになる。これらのＰＣＭ信号サンプルについてさらに１フレームあたり１２ｍｓ～１６ｍｓ程度に分割（フレミング）した（Ｓ４２）後に、後述の音声分析特徴抽出処理により半音節種類×Ｎmax 人分の特徴ベクトル（フレーム単位）を得る（Ｓ４３）。

【００４６】そして、最終的に“Ｋ－”の為の標準パターンを生成するということは、１つの半音節カテゴリー“Ｋ－”に１つのＨＭＭの標準パターンモデルを対応させ、その半音節のＦmax フレーム分の特徴ベクトル出力が対応するＨＭＭの４状態の遷移過程で最も高い確率で出力されるように各状態の確率パラメータおよび状態遷移確率を求めることにある。

【００４７】次に、各半音節種類毎に状態数＝Ｊmax 個のＬｅｆｔ　ｔｏ　Ｒｉｇｈｔ型ＨＭＭの状態出力確率関数Ｂｊの母数とＪmax ×２個分の状態遷移確率＝α（ｊ－１，ｊ）およびα（ｊ，ｊ）の各パラメータを求める方法について説明する。例として、１つの半音節”Ｋ－”に対して、図１１のような状態数が４の半音節ＨＭＭの各パラメータをＮmax 人の特徴ベクトルから求める場合を説明する。

【００４８】ここで、α_{00}～α_{33}は、以下のように状態遷移確率α［＊，＊］を示す。

・α（ｊ－１，ｊ）：状態ｊ－１からｊへの状態遷移確率（α_{01}、α_{02}、α_{03}）。・α（ｊ，ｊ）：状態ｊからｊへの状態遷移確率（α_{00}、α_{11}、α_{22}、α_{33}）。

【００４９】また、出力確率＝B_0～B_3は、以下のような算出式になる。

・Ｂｊ：下式の状態ｊの特徴ベクトル出力確率関数（初期設定する際に多次元正規分布（ガウス）確率密度関数を用いると共に、母数として平均ベクトルおよび共分散行列の各成分を初期値設定する。）

【００５０】

【数１】

13　　14

$$Bj=\sum_{m=0}^{Mmax-1}\mu mj\sum_{k=0}^{Kmax-1}\lambda kmj\cdot(2\pi|Vkmj|)^{-1/2}\cdot Exp[(Xm-\bar{X}kmj)'\cdot Vkmj^{-1}\cdot(Xm-\bar{X}kmj)]$$

（備考：ｔは転置操作（縦ベクトル→横ベクトル）、Ｖｋｍｊ の－１は逆行列を示す。）

ただし、

ｊ＝状態番号、ｊ＝０～Ｊｍａｘ-1

ｋ：混合分布番号。ｋ＝０～Ｋｍａｘ-1、Ｋｍａｘ：混合分布数

ｍ：特徴ベクトル種類番号。ｍ＝０～Ｍｍａｘ-1、Ｍｍａｘ：特徴ベクトル種類数

λｋｍｊ：混合分布の重みを決める混合分布係数

μｍｊ：特徴ベクトル種類間の重みを決める特徴ベクトル重み係数

Ｘｍ：入力音声サンプルのフレーム単位の特徴ベクトル

【００５１】

【数２】

$\bar{X}kmj$：平均ベクトル

Ｖｋｍｊ：共分散行列

｜Ｖｋｍｊ｜：Ｖｍ のノルム（行列式）

なお以降では、説明を容易にするため混合分布数を１（Ｋｍａｘ＝１、λｋｍｊ＝１）、および特徴ベクトル種類を１（Ｍｍａｘ＝１、μｍｊ＝１）として下式を用いる。

［混合分布数＝特徴ベクトル数＝１とした場合の出力確率密度関数］

【００５２】

【数３】

$$Bj(X)=(2\pi|Vj|)^{-1/2}\cdot Exp[(X-\bar{X}j)'\cdot Vj^{-1}\cdot(X-\bar{X}j)]$$

ここで求めるパラメータは、

【００５３】

【数４】

平均ベクトル：$\bar{X}j$

共分散行列：Ｖｊ と、状態遷移確率：α（ｊ－１，ｊ）およびα（ｊ，ｊ）であり、これが半音節”Ｋ－”の標準パターンとなる。

【００５４】これらのパラメータは、Ｎｍａｘ 人分の半音節”Ｋ－”の特徴ベクトルサンプルから以下に述べるＦＢ（Ｆｏｒｗａｒｄ　Ｂａｃｋｗａｒｄ）アルゴリズム（Ｂａｕｍ－ｗｅｌｃｈアルゴリズムともいい、ＥＭ（Ｅｘｐｅｃｔａｔｏｎ　Ｍａｘｉｍｉｚａｔｉｏｎ）手法を基本としたアルゴリズム）により反復的に収束するまで演算を繰り返すことにより得られる。

【００５５】ＦＢアルゴリズムを述べる前に、まず、半音節”Ｋ－”のＮｍａｘ 人分の特徴ベクトルを以下のように再定義する。

【００５６】○再定義

話者ｎの特徴ベクトル：Ｘ→Ｘ（ｎ，ｆ）

ただし、

ｎ：話者番号、ｎ＝０～Ｎｍａｘ-1

ｆ：フレーム番号、ｆ＝０～Ｆｍａｘ(n)-1

Ｆｍａｘ(n)：話者番号ｎの半音節”Ｋ－”のフレーム数

（注：一般に話者毎にサンプルしたフレーム数は異なる）

さらに、以下のＦＢアルゴリズム処理を行う（Ｓ４６）。

【００５７】［ＦＢアルゴリズム］

①共分散行列：Ｖｊ、

【００５８】

【数５】

平均ベクトル：$\bar{X}j$、

状態遷移確率：α［ｊ－１，ｊ］およびα［ｊ，ｊ］の初期値を設定する（Ｓ４５）。

【００５９】［初期設定値］

【００６０】

【数６】

平均ベクトル：$\bar{X}j\rightarrow\bar{X}j\_0$

共分散行列：Ｖｊ→Ｖｊ_０

状態遷移確率：α［ｊ－１，ｊ］→α［ｊ－１，ｊ］_０およびα［ｊ，ｊ］→α［ｊ，ｊ］_０

【００６１】②半音節”Ｋ－”のＨＭＭに対する前向きパスアルゴリズムによる確率値の目標値（＝ＦＷＤ_ｔｈ）と反復処理の最大回数（＝ＣＮＴｍａｘ）を設定する。

【００６２】③　④～⑦の処理をｃｎｔ＝１～ＣＮＴｍａｘまで繰り返す。

【００６３】④　⑤～⑦の処理をｊ＝０～Ｊｍａｘ（Ｊｍａｘ＝３）まで繰り返す。

【００６４】⑤下式により、各パラメータの更新値を算出する。

【００６５】

【数７】

15　　16

$$\alpha[j-1,j]\_cnt=\frac{\sum_{n=0}^{Nmax-1}(1/Nmax)\sum_{f=0}^{Fmax-1}FWD(j-1,f-1)\cdot\alpha[j-1,j]\_(cnt-1)\cdot Bj[X(n,f)]\cdot BCK(j,f)}{\sum_{f=0}^{Fmax-1}FWD(j,f)\cdot BCK(j,f)}$$

$$\alpha[j,j]\_cnt=\frac{\sum_{n=0}^{Nmax-1}(1/Nmax)\cdot\sum_{f=0}^{Fmax-1}FWD(j,f-1)\cdot\alpha[j,j]\_(cnt-1)\cdot Bj[X(n,f)]\cdot BCK(j,f))}{\sum_{f=0}^{Fmax-1}FWD(j,f)\cdot BCK(j,f)}$$

$$\overline{X}j\_cnt=\frac{\sum_{n=0}^{Nmax-1}(1/Nmax)\cdot\sum_{f=0}^{Fmax-1}FWD\_s\cdot Bj[X(n,f)]\cdot BCK(j,f)\cdot X(n,f)}{\sum_{f=0}^{Fmax-1}FWD(j,f)\cdot BCK(j,f)}$$

ただし、

$$FWD\_s=FWD(j-1,f-1)\cdot\alpha[j-1,j]\_(cnt-1)+FWD(j,f-1)\cdot\alpha[j,j]\_(cnt-1)$$

$$Vj\_cnt=\frac{\sum_{n=0}^{Nmax-1}(1/Nmax)\cdot\sum_{f=0}^{Fmax-1}FWD\_s\cdot Bj[X(n,f)]\cdot BCK(j,f)\cdot Vj\_(cnt-1)}{\sum_{f=0}^{Fmax-1}FWD(j,f)\cdot BCK(j,f)}$$

ただし、

$$FWD\_s=FWD(j-1,f-1)\cdot\alpha[j-1,j]\_(cnt-1)+FWD(j,f-1)\cdot\alpha[j,j]\_(cnt-1)$$

$$Vj\_(cnt-1)=[Xm(n,f)-\overline{X}mj\_(cnt-1)]\cdot[Xm(n,f)-\overline{X}mj\_(cnt-1)]t$$

【００６６】⑥新しいパラメータにより、入力特徴ベクトル＝Ｘｍ（ｎ，ｆ）に対する、ＨＭＭモデルの前向きパスアルゴリズムによる出力確率を下式により求める。

【００６７】

【数８】

$$出力確率=(1/Nmax)\cdot\sum_{n=0}^{Nmax-1}FWD(Jmax,Fmax)$$

【００６８】⑦出力確率≧ＦＷＤ＿thが成立するか（Ｓ４７）、あるいは、ｃｎｔ＞ＣＮＴmax となれば（Ｓ４９）処理を終了する。

【００６９】⑧この時の、パラメータを半音節“Ｋ－”の標準パターンとする。

【００７０】ここで、ＦＷＤ（ｊ，ｆ）は、前向きパスアルゴリズムで求められる確率（Ｂａｕｍ－Ｗｅｌｃｈスコアとも呼ばれる）であり、またＢＣＫ（ｊ，ｆ）は、後向きパスアルゴリズムにより求められる確率である。

【００７１】また実際には、標準パターンのパラメータを多数サンプルにより求める際は、状態出力確率関数を２つ以上の多次元正規分布の混合分布としたり（例えば男性と女性別等）、特徴ベクトルの種類を増加させる（例えばメルケプストラムベクトルに加えて、フレーム間差分のΔメルケプストラム、１フレームの平均パワーのフレーム間差分：Δ平均パワー等）ことでより認識精度を向上可能である。

【００７２】以上の処理について、Ｉmax 種類の半音節ＨＭＭを標準パターンとして求めて、標準パターンＲＯＭに格納しておく（Ｓ５０）。

【００７３】（２）音声分析特徴抽出処理

図１２及び図１３は不特定話者半音節音声認識器の機能ブロック構成図である。

【００７４】図１４～図１６は不特定話者半音節音声認識処理のフローチャートである。図１７はフローチャートの凡例を示す図表である。

【００７５】音声分析特徴抽出器の音声分析特徴抽出処理は、以下の全工程をフレーム単位で行う処理である。

【００７６】①入力音声のＰＣＭ信号を１２ｍｓ～１６ｍｓ程度にフレーム分割したフレームＰＣＭ信号をプリエンファシス処理（一次差分処理）することにより高周波数帯域を強調する。

【００７７】②プリエンファシス処理済みフレームＰＣＭ信号に対し窓処理（ハニング窓等）することによりこの後のＦＦＴ処理のためのフレーム境界のスムージング処理をする。

【００７８】③窓処理後のフレームＰＣＭ信号をＮ次ＦＦＴ変換処理することにより線形周波数軸上のＮ次複素係数ベクトルへ変換する。

【００７９】④Ｎ次複素係数ベクトルから複素数の絶対値計算によりＮ次振幅係数ベクトルへ変換する。

【００８０】⑤Ｎ次振幅係数ベクトルに対し対数演算を施してＮ次対数振幅係数ベクトルを算出する。

【００８１】⑥Ｎ次対数振幅係数ベクトルに対し時間軸上への逆離散余弦変換によりＰ次ケプストラムベクトルを算出する。

【００８２】⑦P次ケプストラムベクトルの高時間成分を抑圧する処理（リフタリング）により、声道特性とピッチ特性（声帯特性）を分離し声道特性のみを抽出したP次ケプストラムベクトルを算出する。

【００８３】⑧P次声道特性ケプストラムベクトルに対し周波数軸上への離散余弦変換を行い線形周波数軸上におけるN次声道特性対数振幅係数ベクトルに変換する。

【００８４】⑨N次声道特性対数振幅係数ベクトルに対しメル周波数軸上（近似的に対数スケール）における等分割点上のスペクトル成分へ補間あるいはスムージング処理を施したベクトル成分を算出することにより人間の聴覚周波数分解能特性（低周波：高→高周波：低）に合わせたN次声道特性対数振幅メル尺度係数ベクトルへ変換する。

【００８５】次に、N次声道特性対数振幅メル尺度係数

$$D_{i,j}=\sum_{m=0}^{Nm-1}\sum_{k=0}^{Nk-1}\lambda_{i,j,k,m}\cdot\left\{|V_{i,j,k,m}|+(\Delta X_{i,j,k,m})^{t}\cdot(V_{i,j,k,m})^{-1}\cdot\Delta X_{i,j,k,m}\right\}$$

（備考：tは転置操作（縦ベクトル→横ベクトル）、Vi,j,k,mの－１は逆行列を示す。）

Xin：入力音声の特徴ベクトル

Xi,j,k,m：標準パターンの特徴平均ベクトル

Di,j：半音節＝ｉ、状態＝ｊの状態の状態距離計算値

Vi,j,k,m：標準パターンの特徴量共分散行列

｜Vi,j,k,m｜：共分散行列Vi,j,k,mのノルム（分散値）

ｉ：半音節番号、ｉ＝０～Ｉmax －１、Ｉmax ：全半音節種類

ｊ：１半音節のHMMにおける状態番号、ｊ＝０～Ｊmax －１、Ｊmax ：１HMMの全状態数

ｋ：混合分布番号、ｋ＝０～Kmax －１、Kmax ：混合分布数

ｍ：特徴ベクトル種類番号、ｍ＝０～Mmax －１、Mmax ：総特徴ベクトル種類数

【００９０】（４）Ｖｉｔｅｒｂｉ処理（パターンマッチング処理）

例として、以下のような氏名リストの認識を行うことを考える。“たかはし”という単語はまず、“ＴＡＫＡＨＡＳＩ”と母音、子音列に変換され、更に“Ｔ－，－Ａ－，－Ｋ－，－Ａ－，－Ｈ－，－Ａ－，－Ｓ－，－Ｉ”という半音節列に規則的に、分解される。

（単音節列）“たかはし”

（母音、子音列）“ＴＡＫＡＨＡＳＩ”

（半音節列）

“Ｔ－，－Ａ－，－Ｋ－，－Ａ－，－Ｈ－，－Ａ－，－Ｓ－，－Ｉ”

各半音節は、前述の図１１のように標準パターンHMMを持って表現されていた。

【００９１】これにより、“たかはし”という単語のHMM連結モデルは図１８のようになる。

【００９２】このHMM連結モデルから、一種の累積確

ベクトルに対し時間軸上への逆離散余弦変換によりＱ次メルケプストラムベクトルへ変換する。

【００８６】以上により、１フレーム分ＰＣＭ信号から入力特徴ベクトルが得られる（Ｓ５８）。

【００８７】（３）状態距離計算

状態距離計算器は、入力音声のアナログ信号をＡＤ変換し、さらに１２ｍｓ～１６ｍｓ程度にフレーム分割したＰＣＭ信号から音声分析特徴抽出器の音声分析特徴抽出処理により得られた入力特徴ベクトルと、標準半音節HMMの全状態の標準特徴ベクトルとの状態距離値を下記の距離計算式により算出する（Ｓ６０）。

【００８８】［距離計算式］

$\Delta X_{i,j,k,m}=X_{in}-X_{i,j,k,m}$

【００８９】

【数９】

率を算出するのがＶｉｔｅｒｂｉアルゴリズムである。

【００９３】Ｖｉｔｅｒｂｉアルゴリズムは、基本的に図１９の最適パス選択処理の繰り返しである。

【００９４】まず分かりやすい例としてとして、１つの半音節HMM“Ｔ－”のＶｉｔｅｒｂｉスコア算出例を図２０に示す。又、入力パターンはＦmaxフレーム分とする。　図２０の例のように、各フレーム入力毎に、全状態のＶｉｔｅｒｂｉスコアを求めていき、全フレーム分について、算出した時の状態３のＶｉｔｅｒｂｉスコアが、“Ｔ－”の入力特徴ベクトルに対する標準パターンの出力確率となる。さて、これを“たかはし”という単語のHMM連結モデルに適用する場合には、状態数が、４×８＝３２、入力フレーム数＝Ｆmax として、半音節“－Ｉ”のHMMにおける状態３のＶｉｔｅｒｂｉスコアを算出することで、入力特徴ベクトル“たかはし”のHMM連結モデルからの出力確率が求まることになる。

【００９５】実際には、認識対象リストが、“たかはし”の他にも複数存在するので、例えば“いとう”という氏名に対しても同様の半音節列への変換をしてＶｉｔｅｒｂｉスコアを算出する（Ｓ７５）。そして、全認識対象リストにおけるＶｉｔｅｒｂｉスコアから最も確率値の高い（距離値の小さい）認識対象リストの１つを認識結果とする（Ｓ８６）。以下にＶｉｔｅｒｂｉアルゴリズムの処理手順を示す。

【００９６】［Ｖｉｔｅｒｂｉアルゴリズム］

①（Ｓ５７～Ｓ８１）②～⑥の処理をｆ＝０～Ｆmax －１まで繰り返す。

②（Ｓ７３～Ｓ７９）③～⑥の処理をｗ＝０～Ｗmax －１まで繰り返す。

③（Ｓ７３～Ｓ７７）④～⑥の処理をｓ＝０～Ｓmax（ｗ）－１　まで繰り返す。

④（Ｓ７３）ｉ←［状態＝ｓが属する半音節番号］

ｊ←［状態＝ｓが属している半音節番号＝ｉにおけるＨＭＭ内の状態番号］

⑤（Ｓ７４）Ｐａｔｈ（ｊ－１，ｊ）＝α（ｊ－１，ｊ）＋Ｇ（ｗ，ｓ－１）

Ｐａｔｈ（ｊ，ｊ）＝α（ｊ，ｊ）＋Ｇ（ｗ，ｓ）

⑥（Ｓ７５）累積距離値：Ｇ（ｗ，ｓ）＝Ｍａｘ［Ｐａｔｈ（ｊ－１，ｊ），Ｐａｔｈ（ｊ，ｊ）］＋Ｄｉｊの計算

ただし、

ｆ：入力フレーム番号、Ｆｍａｘ：全フレーム数

ｗ：認識対象（氏名ｏｒ電話番号）リストの番号、Ｗｍａｘ：全リスト数

ｓ：リスト内状態連結の通し番号、Ｓｍａｘ（ｗ）：認識対象リスト＝ｗ番の全連結状態数

α（ｊ－１，ｊ）：ｉ番の半音節ＨＭＭにおいて状態ｊ－１からｊへの状態遷移距離値（ｊ－１＜０の場合は距離値＝０）

α（ｊ，ｊ）：ｉ番の半音節ＨＭＭにおいて状態ｊからｊへの状態遷移距離値

Ｄｉｊ：半音節＝ｉ、状態＝ｊの状態の入力特徴ベクトルとの状態距離値

【００９７】なお、Ｖｉｔｅｒｂｉスコアは確率値として説明していたが、実際には桁数の制限等でアンダーフロー等の問題を避ける為に、対数演算を施したものでＶｉｔｅｒｂｉスコアを計算することもある。

【００９８】又、演算量低減及びメモリ量低減の為に標準パターンの平均ベクトル、共分散行列をクラスタリングして演算量を低減する工夫もある。

【００９９】例えば、半音節ＨＭＭが２５０種類あると、２５０×４＝１０００種類の平均ベクトルと共分散行列を用意しなければならないが、例えば平均ベクトルを５１２カテゴリー（この場合、共分散行列も５１２種類）に、また分散行列のノルムを２５６カテゴリーに、ベクトル量子化の手法（セントロイドベクトル等）によりベクトル値を代表させることで、クラスタリングを行うと、演算量とメモリ量が１／２～１／４になる。実際にこのような工夫により、認識性能は劣化することなく演算量とメモリ量を低減することは可能である。

【０１００】４．音声登録用検索器

音声登録用検索器（以降検索器という）は、不特定話者半音節認識器（以降認識器という）から１音節単位の認識結果を第１候補～Ｎ候補まで距離値と共に情報受信し、データベースＲＯＭ１に格納されている日本人全氏名の平仮名文字データリスト内で、１音節単位に発声された氏名の音声信号から最も距離の近い氏名の文字コードを検索し、これを認識音節列ワークＲＡＭへ格納する。これにより、音声信号入力による氏名登録が行われることになる。以降では、この音声登録動作について具体的な実施例を説明する。

【０１０１】まず、発声される音声信号を“た”＋

“か”＋“は”＋“し”と、１音節単位に離散発声されたものを例とする。認識器から検索器には、１音節の発声毎に認識結果の文字コードと距離値が図２０の例のように複数候補出力される。

【０１０２】ここで、第１候補が実際の発声音声＝“た”に対し“な”になっているのは、１音節分の発声方法が似かよっているためである。これは、基本的に子音＋母音という１音節の構造上子音だけが異なり、母音が全て同じ場合には、たとえ人の聴覚識別能力であっても間違って聞き取ってしまうものと同じ事である。認識器では、メルケプストラムという音声信号の特徴量を抽出しているがこれは声道の特徴量を抽出していることと同じ意味である。

【０１０３】“た”のＰＣＭ信号を実際に認識器により認識処理した場合した場合は、“Ｔ－”＋“－Ａ”という半音節ＨＭＭ結合より出現される確率が最も高い（距離値が小さい）のではなく、実際には人間の微妙な発声変形等の影響により、“子音の半音節”＋“－Ａ”もほぼ同等の認識距離となってしまうのである。例えば同母音系の”な”を考えてみると、半音節は“Ｎ”＋“－Ａ”となり、母音部は同じとなるため、また子音部についても”た”と”な”については、どちらも子音を発声する瞬間は舌を上あごにつけてから“－Ａ”を発声するため、“た”と誤認識し易くなってしまうのである。しかし、複数音節から通常構成される氏名の場合は、上例において残る“か”、“は”、“し”の音節の認識結果も誤認識し易くなるのは変わりないが、累積確率を存在する氏名の音節列のみに対して計算することにより、単音節認識の誤認識を補うことが出来るのである。これを以降に示す。

【０１０４】まず、“た、”、“か”、“は”、“し”の各発声に対する単音節認識結果が図２２の例のようになったとする。

【０１０５】まず、検索装置ではこの情報を登録用音節列ワークＲＡＭへ一時格納しておく。次に、発声回数が４回であることがカウンタ等により力ウント出来るため、データベースＲＯＭから四文字の氏名リストを検索し、そられもワークＲＡＭに格納しておく。次に、４文字氏名リストの文字コードを調べて認識結果の全ての文字コードが１つでも含まれている氏名を絞り込む。更に、絞り込んだ４文字氏名リストについて認識結果の距離値でもって累積距離値を累加算演算処理により計算していく。ここで、ある氏名のある文字コードには距離がないものがあるが、以下の方法により計算する。

【０１０６】ここで、また例として絞り込まれた４文字氏名リストが図２３のようになったとする。

【０１０７】次に、これらの候補氏名に対して累積距離値を算出する。この際認識結果リストにある候補に対しては、発声順も考慮して距離値を加算していき、認識結果リストにない文字に対しては最大距離値＝５．０を設

定して累加算演算を行う。図２３の例の一覧により、“た”＋“か”＋“は”＋“し”が最も小さい値になり、これが登録氏名となる。

【０１０８】

【発明の効果】以上説明したように本発明は、入力音声に対する、登録済み氏名又は電話番号の単音節コードデータ列と半音節単位の標準パターンデータとを用いた認識処理により、携帯電話機等における氏名と電話番号登録操作、およびダイアル操作等のキー操作における煩雑さを低減することができ、音声認識ダイアル用の氏名と電話番号の登録操作時間が従来例と比べて短縮され、また操作間違いも少なくなるという効果がある。

【０１０９】また、従来技術の認識自動ダイアル電話機では、明記されていなかった音声から文字表示を行う音声認識処理について、この技術の実現手段を明記することで本発明の属する技術分野において実際に利用可能となるという効果がある。

【図面の簡単な説明】

【図１】本発明の実施の形態の全体構成を示すブロック図である。

【図２】音声登録動作モードのフローチャートである。

【図３】不特定話者半音節音声認識器により得られた４音節×５候補分の認識結果を示す図である。

【図４】電話番号登録動作モードのフローチャートである。

【図５】電話番号の認識結果の数字を直接表すコードへの変換を示す図である。

【図６】音声ダイアル動作モードのフローチャートである。

【図７】標準パターン生成器の機能ブロック構成図である。

【図８】標準パターン生成器の機能ブロック構成図である。

【図９】標準パターン生成処理のフローチャートである。

【図１０】単音節ＰＣＭ信号から半音節ＰＣＭ信号を得る例の図である。

【図１１】状態数が４の半音節ＨＭＭを示す図である。

【図１２】不特定話者半音節音声認識器の機能ブロック構成図である。

【図１３】不特定話者半音節音声認識器の機能ブロック構成図である。

【図１４】不特定話者半音節音声認識処理のフローチャートである。

【図１５】不特定話者半音節音声認識処理のフローチャートである。

【図１６】不特定話者半音節音声認識処理のフローチャートである。

【図１７】フローチャートの凡例を示す図表である。

【図１８】“たかはし”という単語のＨＭＭ連結モデルを示す図である。

【図１９】最適パス選択処理を示す図である。

【図２０】１つの半音節ＨＭＭ“Ｔ－”のＶｉｔｅｒｂｉスコア算出例を示す図である。

【図２１】不特定話者半音節音声認識器から音声登録用検索器に複数候補出力される、１音節の発声毎の認識結果の文字コードと距離値を示す図である。

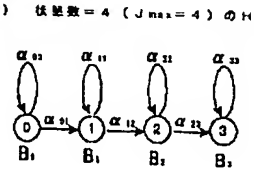
【図２２】“た、”、“か”、“は”、“し”の各発声に対する単音節認識結果を示す図である。

【図２３】絞り込まれた４文字氏名リストに対して累積距離値を算出する図である。

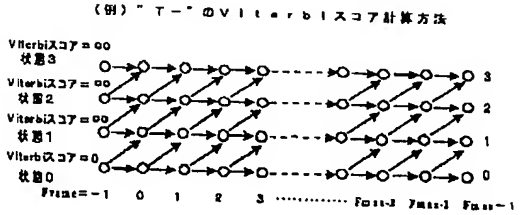
【符号の説明】

１	データベースＲＯＭ
２	標準パターンＲＯＭ
３	認識ワークＲＡＭ
４	検索ワークＲＡＭ
５	標準パターン生成器
６	不特定話者半音節音声認識器
７	音声登録用検索器
８	ＡＤコンバータ
９	ＤＡコンバータ
１０	音声合成器
１１	携帯電話本体部
１２	キーマトリクス
１３	表示装置
１４	Ｉ／Ｆ部
１５	音声登録音声ダイアル操作者
１６	制御信号バス
１７	メモリバス

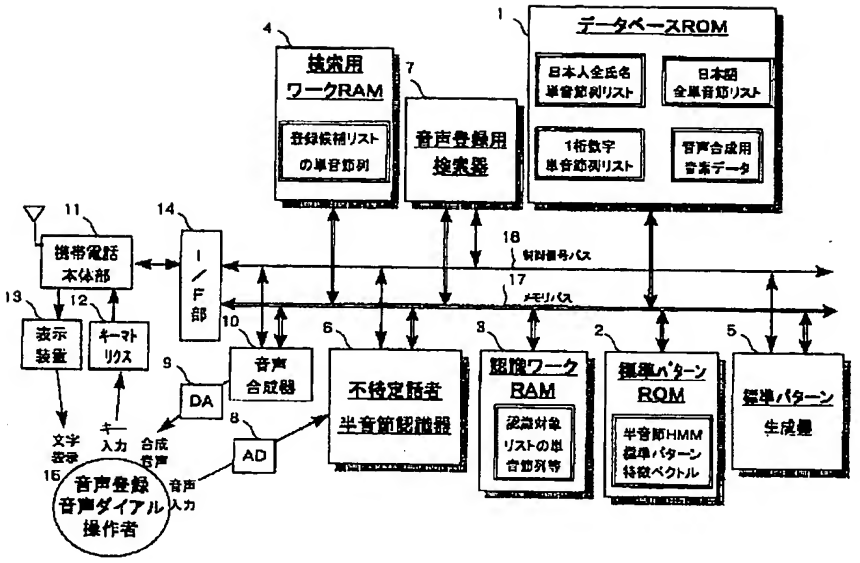
【図１１】



【図２０】



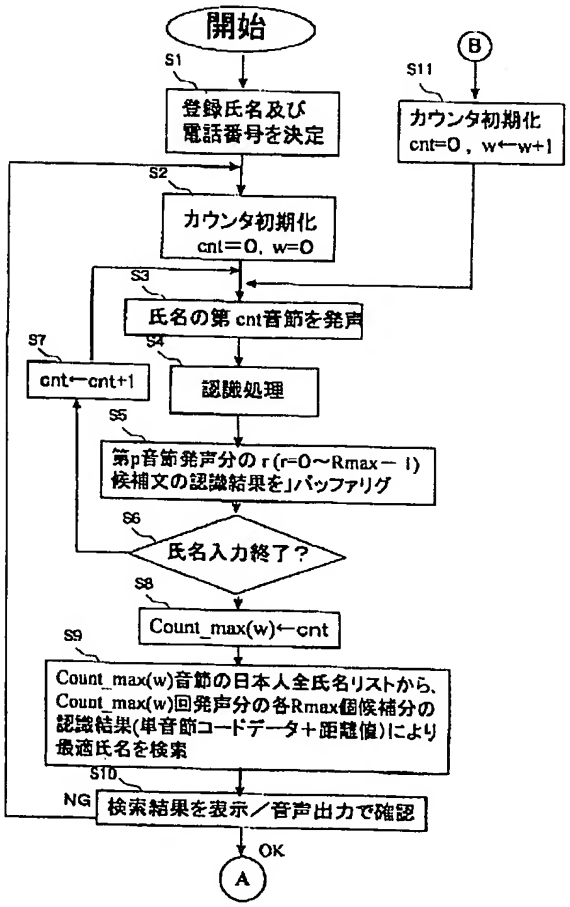
【図１】



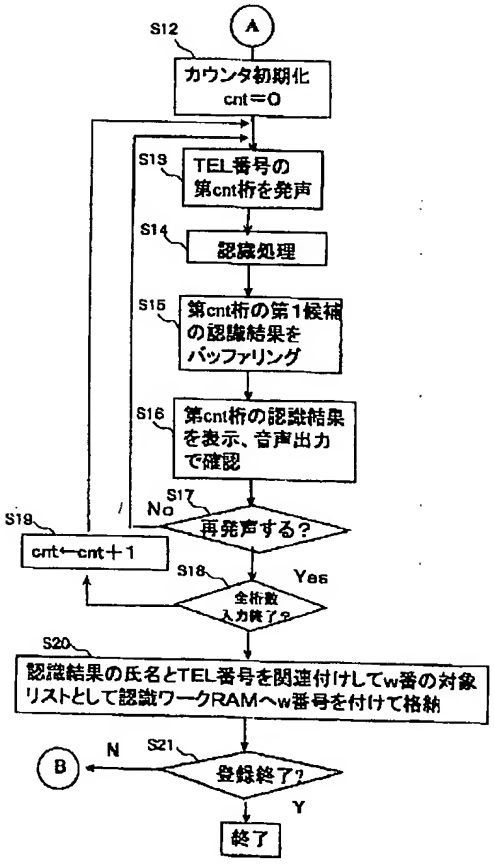
【図５】

認識結果	数字表記
"ぜ＋ろ"	→ "０"
"い＋ち"	→ "１"
"　に　"	→ "２"
"さ＋ん"	→ "３"
"よ＋ん"	→ "４"
"　ご　"	→ "５"
"ろ＋く"	→ "６"
"な＋な"	→ "７"
"は＋ち"	→ "８"
"き＋ゅ＋う"	→ "９"

【図２】



【図４】



【図３】

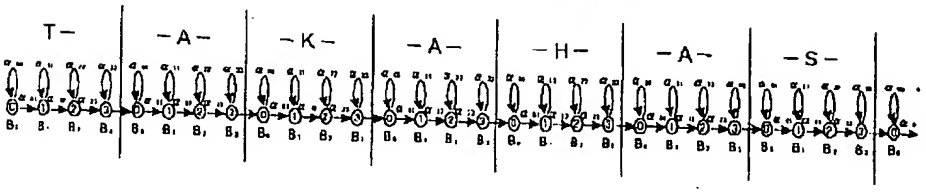
（例）

	第１候補	第２候補	第３候補	第４候補	第５候補
第１発声：“た”	な	た	ら	さ	か
	(1.5)	(1.7)	(2.6)	(3.5)	(4.5)
第２発声：“か”	あ	は	か	や	わ
	(1.5)	(1.7)	(1.8)	(3.5)	(4.5)
第３発声：“は”	あ	は	か	や	わ
	(1.4)	(1.5)	(1.7)	(3.0)	(4.5)
第４発声：“し”	い	し	ひ	ち	じ（ぢ）
	(1.4)	(1.5)	(1.6)	(2.0)	(3.0)

（注）括弧内の数値は不特定話者半音節音声認識器から出力される距離値。

【図１８】

（例）　“たかはし”のＨＭＭ連結モデル



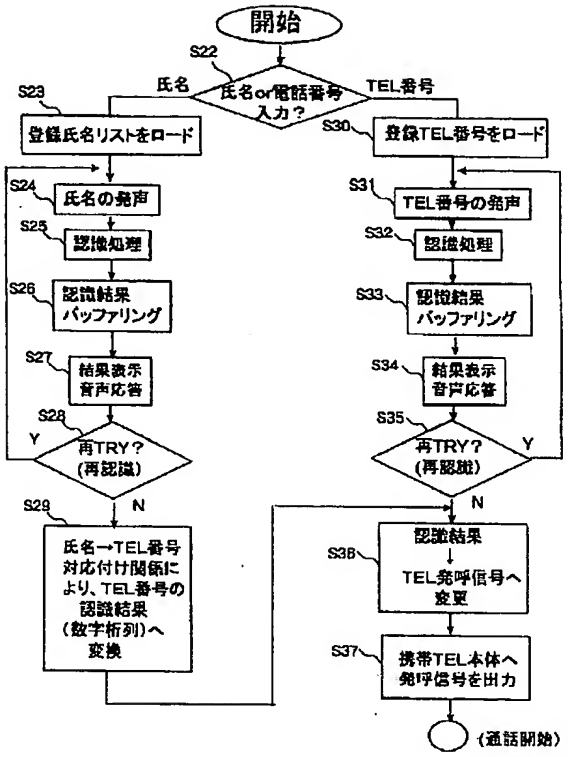
【図２１】

（例）第１発声＝“た”に対する１音節認識結果

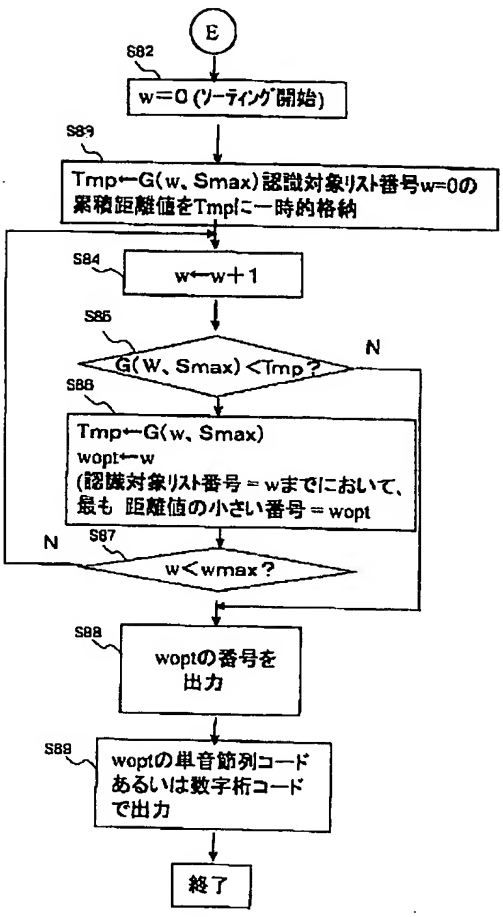
候補	第１位	第２位	第３位	第４位	第５位
音節	な	た	ら	さ	か
距離値	１．５	１．７	２．６	３．５	４．５

（注）実際は、音節は音節コード値。距離値範囲：1.0～5.0。

【図６】

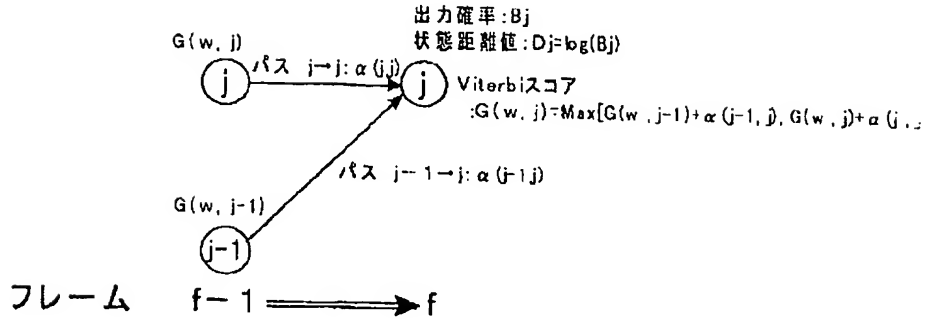


【図１６】

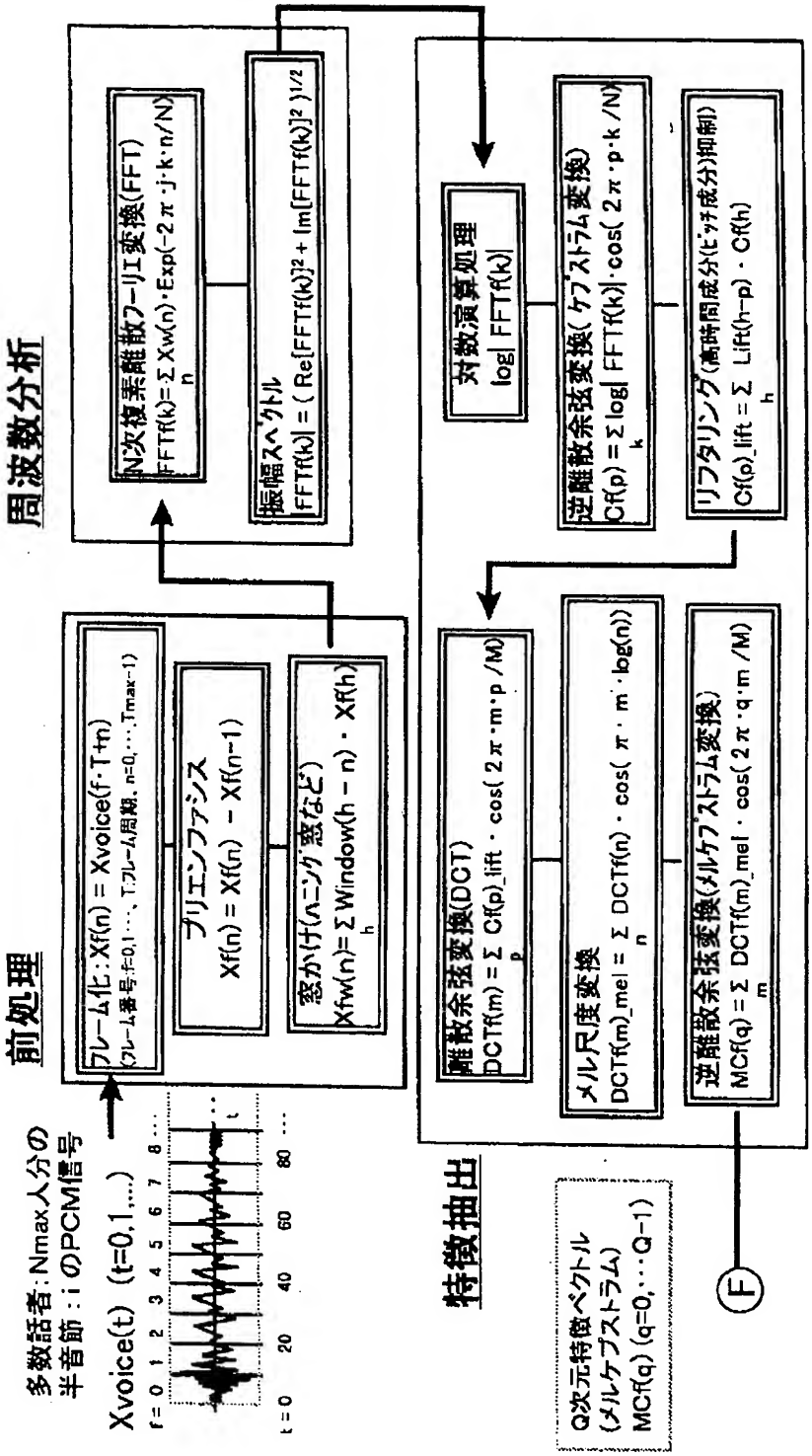


【図１９】

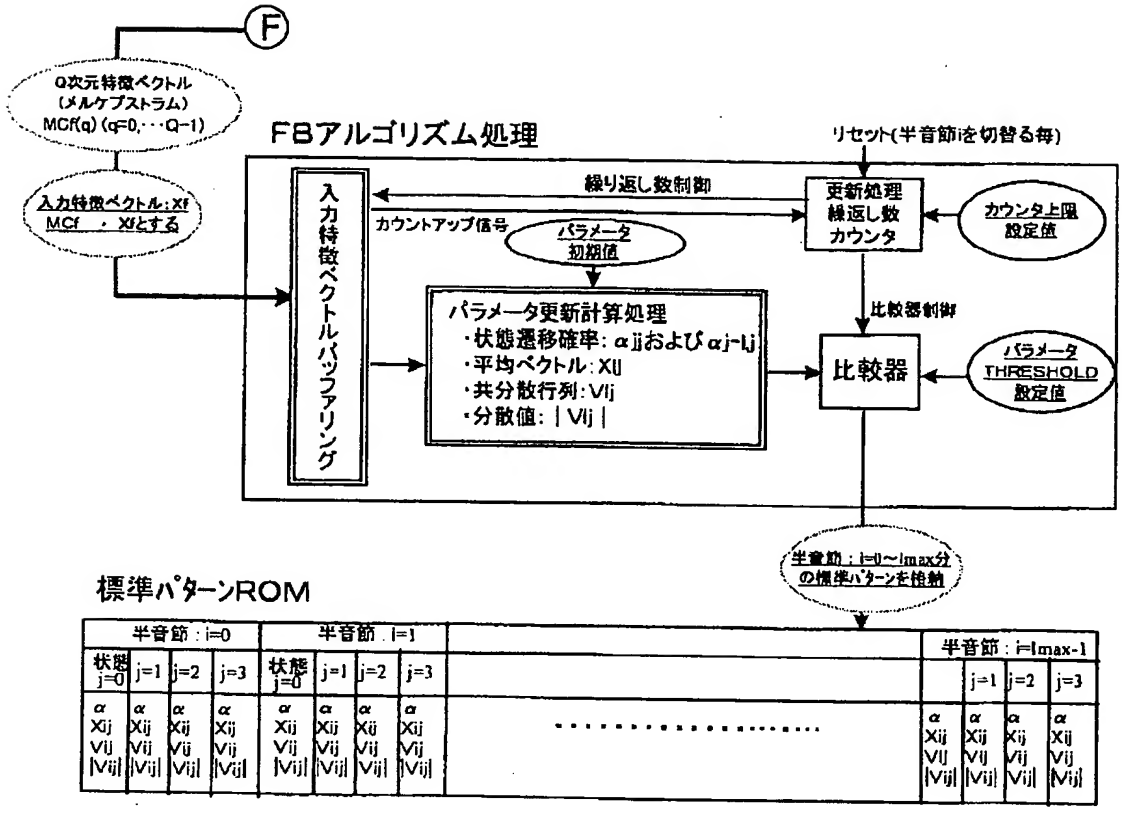
（例）最適パスの選択方法



【図７】


多数話者：Nmax人分の
半音節：iのPCM信号
Xvoice(t)　(t=0,1,...)
前処理
フレーム化：Xf(n) = Xvoice(f･T+n)
プリエンファシス
Xf(n) = Xf(n) − Xf(n-1)
窓かけ(ハニング窓など)
Xfw(n)= Σ Window(h − n) ･ Xf(h)
周波数分析
N次複素離散フーリエ変換(FFT)
FFTf(k)= Σ Xw(n)･Exp(−2π･j･k･n/N)
振幅スペクトル
FFTf(k)| = (Re[FFTf(k)]² + Im[FFTf(k)]²)1/2
対数演算処理
log| FFTf(k)|
逆離散余弦変換(ケプストラム変換)
Cf(p) = Σ log| FFTf(k)|･cos(2π･p･k /N)
リフタリング(高時間成分(ピッチ成分)抑制)
Cf(p)_lift = Σ Lift(h−p) ･ Cf(h)
離散余弦変換(DCT)
DCTf(m) = Σ Cf(p)_lift ･ cos(2π･m･p /M)
メル尺度変換
DCTf(m)_mel = Σ DCTf(n) ･ cos(π ･ m ･log(n))
逆離散余弦変換(メルケプストラム変換)
MCf(q) = Σ DCTf(m)_mel ･ cos(2π･q･m /M)
特徴抽出
Q次元特徴ベクトル
(メルケプストラム)
MCf(q) (q=0,･･･Q−1)
F

【図８】



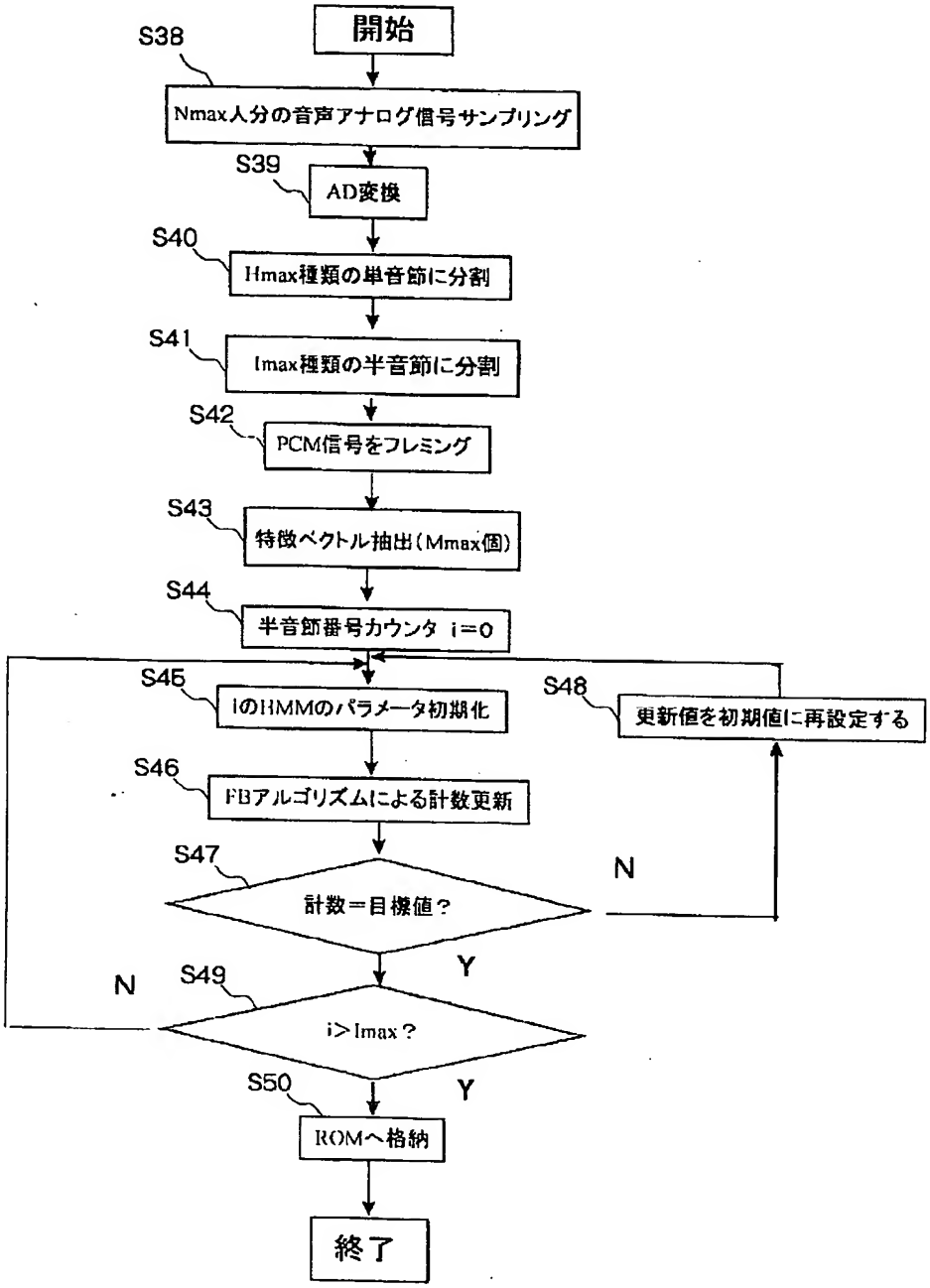
(注) f:フレーム番号, i:半音節番号, j:状態番号　平均ベクトル:Xij、共分散行列:Vij、分散値:|Vij|

【図２２】

(例)

	第１候補	2	3	4	5
発声 “た”	な (1.5)	た (1.7)	ら (2.6)	さ (3.5)	か (4.5)
“か”	あ (1.5)	は (1.7)	か (1.8)	や (3.5)	わ (4.5)
“は”	あ (1.4)	は (1.5)	か (1.7)	や (3.0)	わ (4.5)
“し”	い (1.4)	し (1.5)	ひ (1.6)	ち (2.0)	じ(ぢ) (3.0)

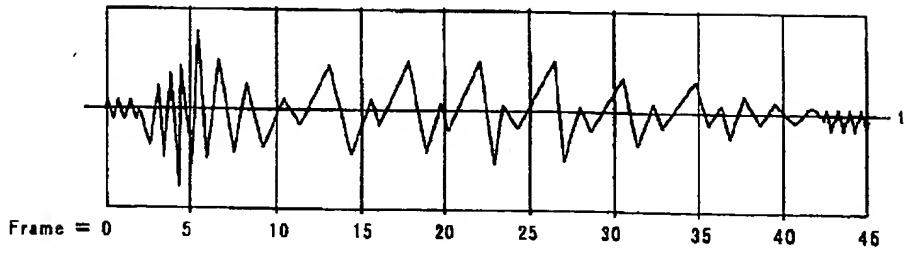
【図９】


開始
S38
Nmax人分の音声アナログ信号サンプリング
S39
AD変換
S40
Hmax種類の単音節に分割
S41
Imax種類の半音節に分割
S42
PCM信号をフレミング
S43
特徴ベクトル抽出(Mmax個)
S44
半音節番号カウンタ i=0
S45
iのHMMのパラメータ初期化
S46
FBアルゴリズムによる計数更新
S47
計数＝目標値？
N
Y
S48
更新値を初期値に再設定する
S49
i>Imax？
N
Y
S50
ROMへ格納
終了

【図１０】

(例)ある話者ｎの発声音：単音節"ＫＡ"のＰＣＭ信号波形

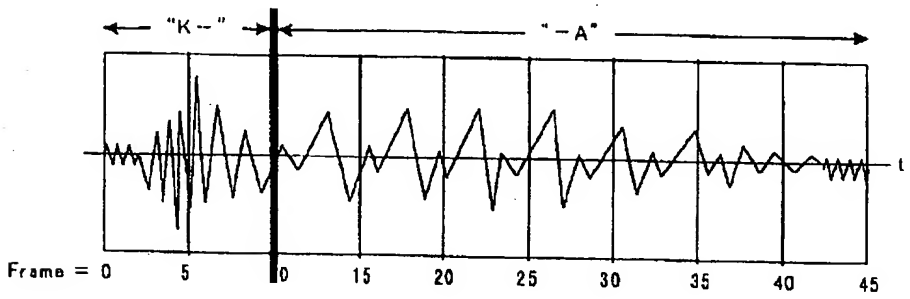
[単音節"ＫＡ"(か)のＰＣＭ信号波形]



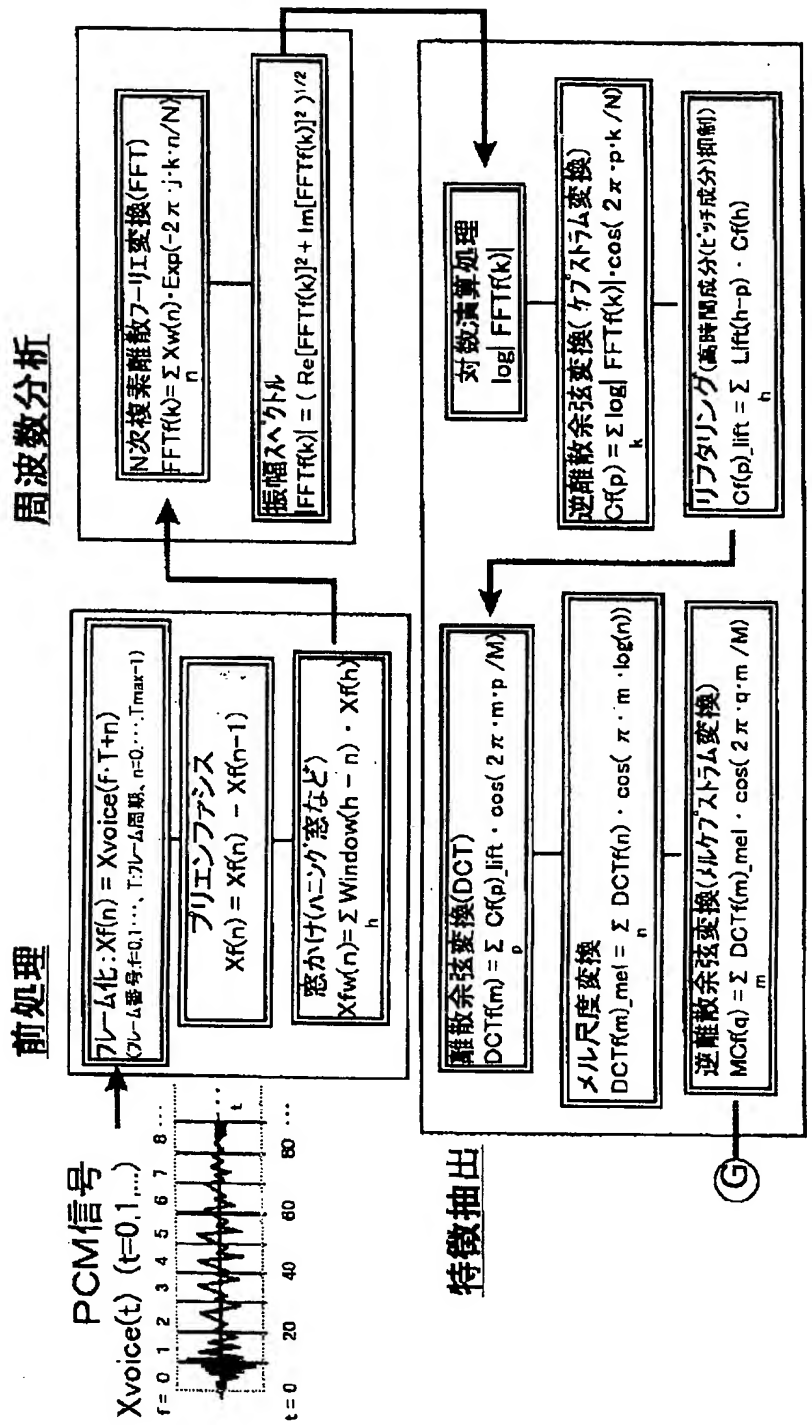
(1Frame：12～16ms)

↓

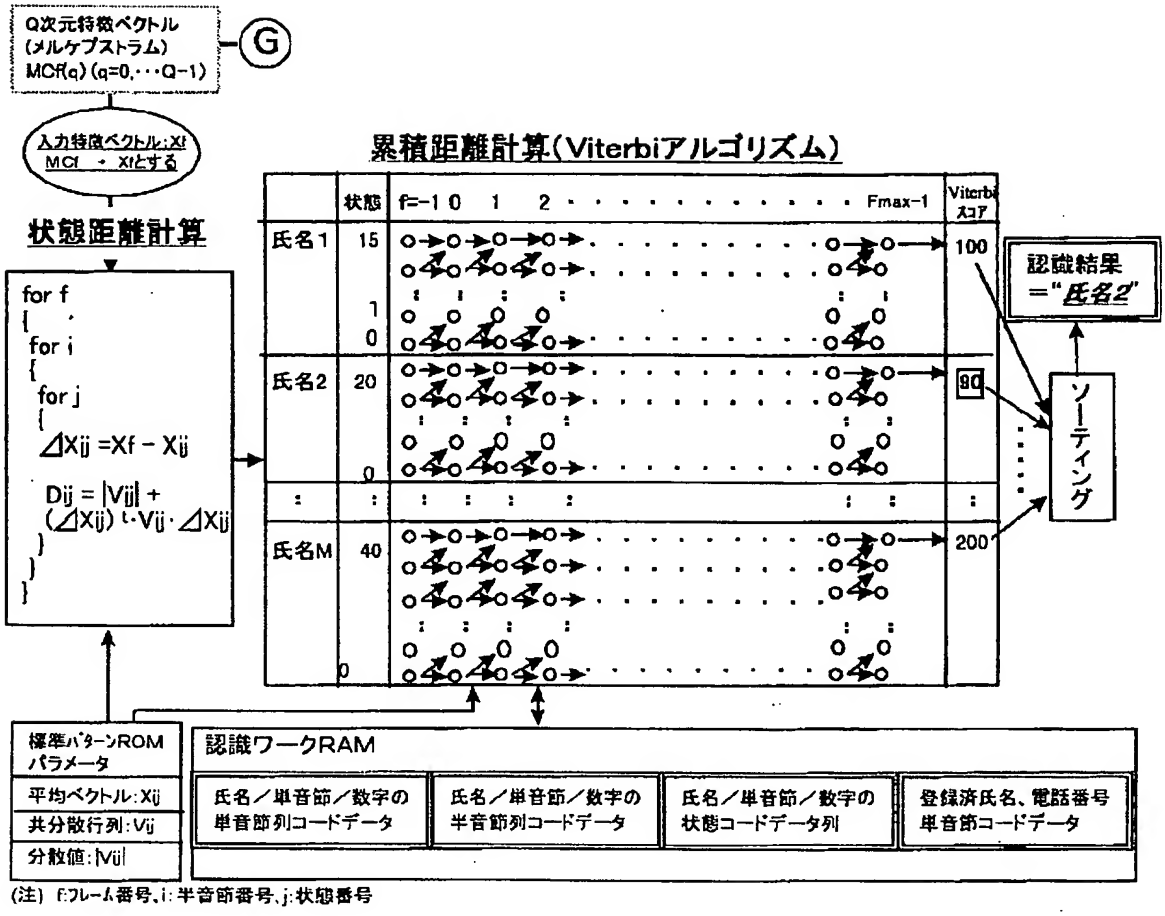
[半音節"Ｋ－"＋"－Ａ"に分割されたＰＣＭ信号波形]



【図１２】

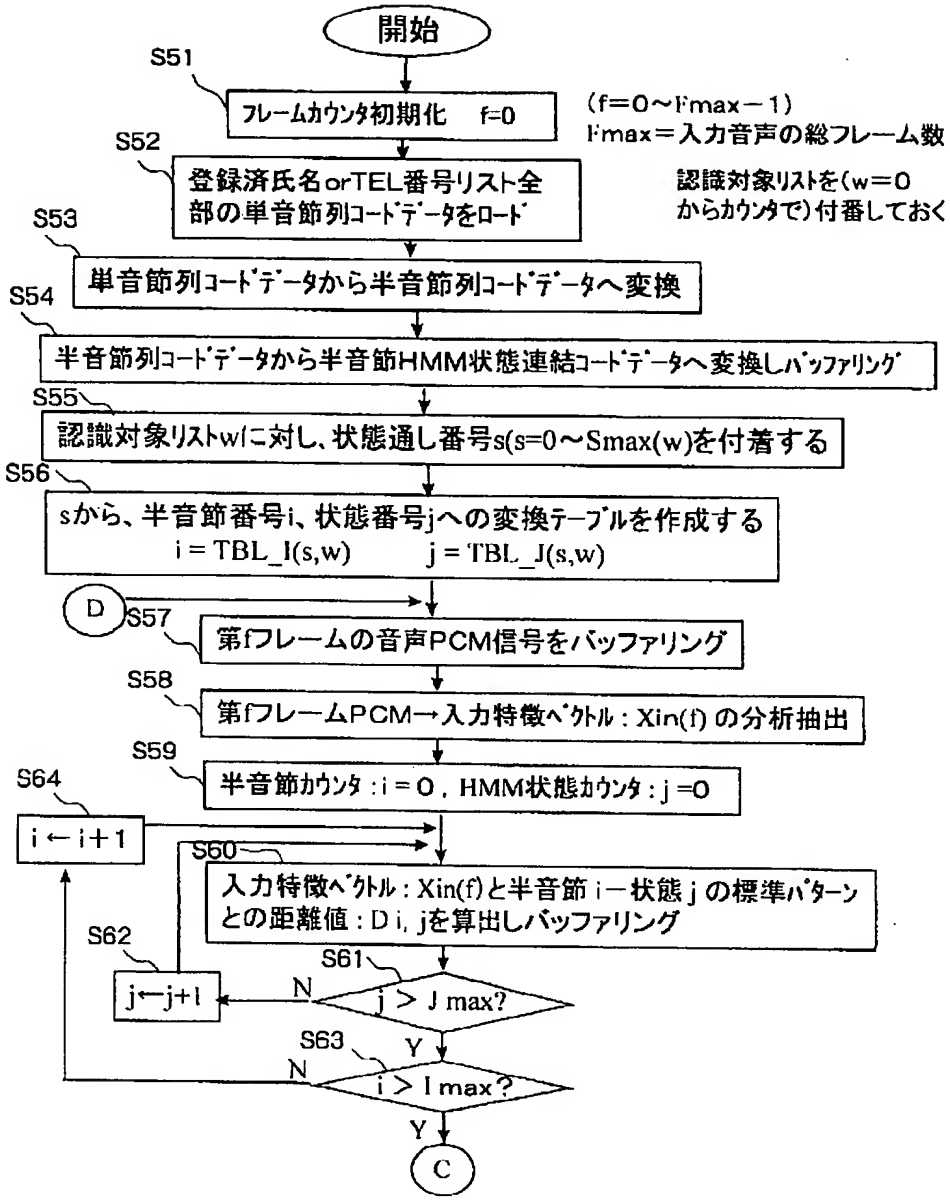

前処理
PCM信号
Xvoice(t) (t=0,1,...)
フレーム化：Xf(n) = Xvoice(f･T+n)
プリエンファシス
Xf(n) = Xf(n) − Xf(n−1)
窓かけ(ハニング窓など)
Xfw(n)= Σ Window(h − n) ・ Xf(h)
周波数分析
N次複素離散フーリエ変換(FFT)
振幅スペクトル
|FFTf(k)| = (Re[FFTf(k)]² + Im[FFTf(k)]²)^1/2
対数演算処理
log| FFTf(k)|
逆離散余弦変換(ケプストラム変換)
リフタリング(高時間成分(ピッチ成分)抑制)
特徴抽出
離散余弦変換(DCT)
メル尺度変換
逆離散余弦変換(メルケプストラム変換)

【図１３】

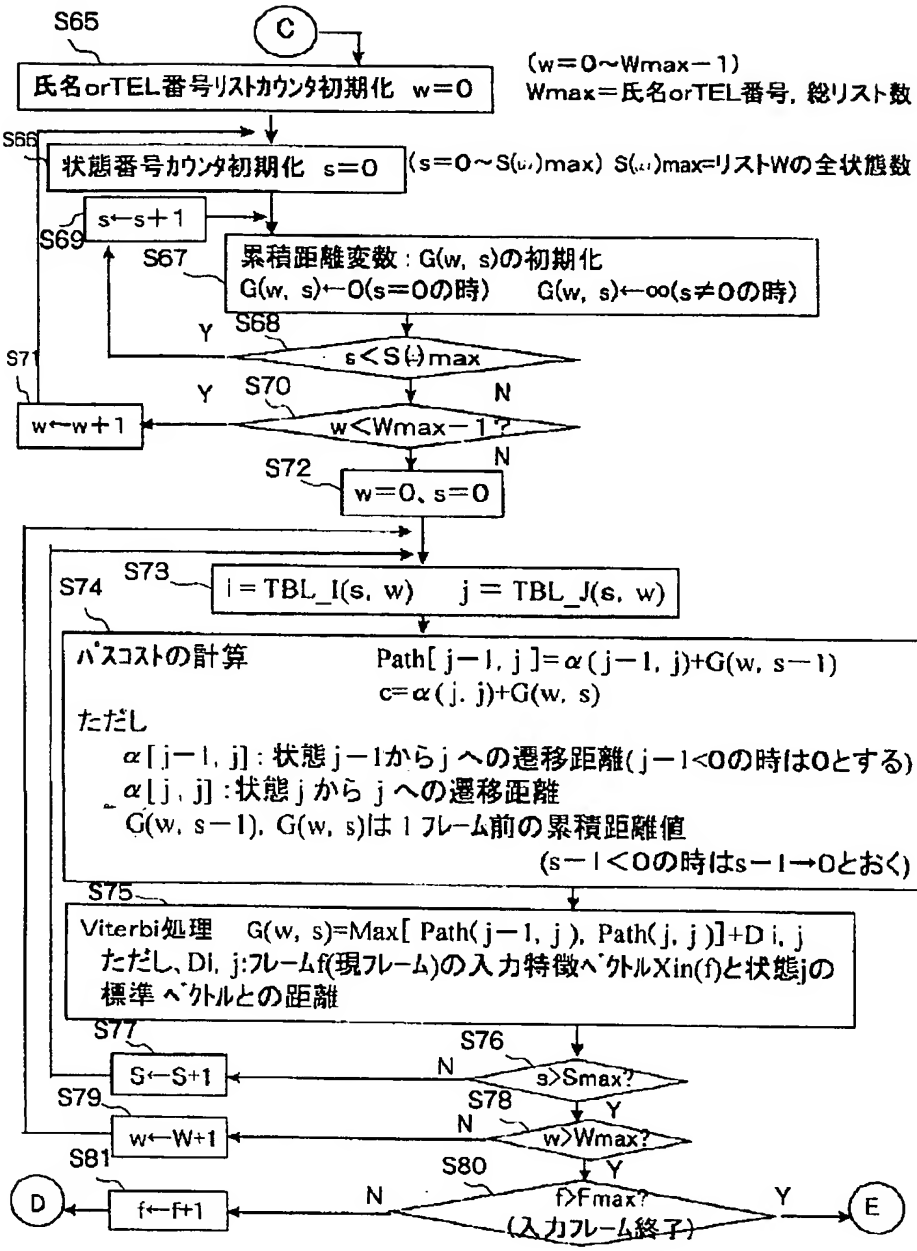


(注) f:フレーム番号、i:半音節番号、j:状態番号

【図１４】


開始
S51
フレームカウンタ初期化　f=0
(f=0～Fmax－1)
Fmax＝入力音声の総フレーム数
S52
登録済氏名orTEL番号リスト全部の単音節列コードデータをロード
認識対象リストを(w＝0からカウンタで)付番しておく
S53
単音節列コードデータから半音節列コードデータへ変換
S54
半音節列コードデータから半音節HMM状態連結コードデータへ変換しバッファリング
S55
認識対象リストwに対し、状態通し番号s(s=0～Smax(w)を付着する
S56
sから、半音節番号i、状態番号jへの変換テーブルを作成する
i = TBL_I(s,w)　　j = TBL_J(s,w)
D
S57
第fフレームの音声PCM信号をバッファリング
S58
第fフレームPCM→入力特徴ベクトル：Xin(f) の分析抽出
S59
半音節カウンタ：i＝0，HMM状態カウンタ：j＝0
S64
i←i+1
S60
入力特徴ベクトル：Xin(f)と半音節 i－状態 j の標準パターンとの距離値：D i, jを算出しバッファリング
S62
j←j+1
S61
N
j > J max?
Y
S63
N
i > I max?
Y
C

【図１５】


C
S65
氏名orTEL番号リストカウンタ初期化 w=0
(w=0～Wmax－1)
Wmax=氏名orTEL番号，総リスト数
S66
状態番号カウンタ初期化 s=0
(s=0～S(w)max) S(w)max=リストWの全状態数
S69
s←s+1
S67
累積距離変数：G(w, s)の初期化
G(w, s)←0(s=0の時) G(w, s)←∞(s≠0の時)
Y
S68
s<S(w)max
N
Y
S70
w<Wmax－1?
N
S71
w←w+1
S72
w=0、s=0
S73
i = TBL_I(s, w) j = TBL_J(s, w)
S74
パスコストの計算 Path[j－1, j]=α(j－1, j)+G(w, s－1)
c=α(j, j)+G(w, s)
ただし
α[j－1, j] : 状態 j－1から j への遷移距離(j－1<0の時は0とする)
α[j , j] : 状態 j から j への遷移距離
G(w, s－1), G(w, s)は 1 フレーム前の累積距離値
(s－1<0の時はs－1→0とおく)
S75
Viterbi処理 G(w, s)=Max[Path(j－1, j), Path(j, j)]+D i, j
ただし、Di, j:フレームf(現フレーム)の入力特徴ベクトルXin(f)と状態jの標準ベクトルとの距離
S77
S←S+1
S76
N
s>Smax?
S78
Y
S79
w←W+1
N
w>Wmax?
Y
S81
S80
D
f←f+1
N
f>Fmax?
(入力フレーム終了)
Y
E

【図１７】

	Index	値域
標準話者数	n	$0 \leqq n \leqq Nmax-1$
単音節番号	h	$0 \leqq h \leqq Hmax-1$
半音節番号	i	$0 \leqq i \leqq Imax-1$
状態番号 （半音節番号 i 内の状態番号）	j	$0 \leqq j \leqq Jmax-1$
認識対象リスト番号 （氏名あるいはTELリストなど）	w	$0 \leqq w \leqq Wmax-1$
フレーム番号	f	$0 \leqq f \leqq Fmax-1$
サンプリング時間	t	$-\infty \leqq t \leqq \infty$
連結状態番号 （リストw内の連結状態の通し番号）	s	$0 \leqq s \leqq Smax(w)-1$
混合分布	k	$0 \leqq k \leqq Kmax-1$
特徴ベクトル種類	m	$0 \leqq m \leqq Mmax-1$
認識結果候補カウンタ	r	$0 \leqq r \leqq Rmax$
汎用カウンタ	cnt	$0 \leqq cnt \leqq \infty$

【図２３】

(例)	第１発声	2	3	4　累積距離値
1	"あ" (５)	"か" (1.8)	"ざ" (５)	"わ" (５) →１６．８
2	"か" (4.5)	"き" (５)	"さ" (５)	"わ" (５) →１９．５
3	"さ" (3.5)	"わ" (4.5)	"は" (1.5)	"し" (1.5) →１１．０
4	"た" (1.7)	"き" (５)	"ざ" (５)	"わ" (５) →１６．７
5	"た" (1.7)	"か" (1.8)	"は" (1.5)	"し" (1.5) → 6．5
6	"あ" (1.5)	"か" (1.8)	"ざ" (５)	"わ" (５) →１３．３
7	"は" (1.5)	"し" (1.8)	"か" (５)	"わ" (５) →１３．３
8	"ま" (1.5)	"え" (1.8)	"か" (５)	"わ" (５) →１３．３
9	"み" (1.5)	"つ" (1.8)	"は" (５)	"し" (５) →１３．３
１０	"や" (1.5)	"ま" (1.8)	"か" (５)	"わ" (５) →１３．３